

Title: US-10-046-433-40
 Perfect score: 5506
 Sequence: MAEFGHSHLSARVGRTER.....LGRSNHPPRGLIMDLTQCR 1001

Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5
 112892 seqs, 41476328 residues

Searched: Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
 Maximum DB seq length: 200000000

post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Copyright (c) 1993 - 2003 Compugen Ltd.

GenCore version 5.1.4_p5-4578

OM protein - protein search, using sw model

Run on: March 11, 2003, 23:10:30 ; Search time 23 seconds
 (without alignments)
 1805.122 Million cell updates/sec

RESULT 1

ID	LMGL_HUMAN	STANDARD:	PRT;	1609 AA.
AC	P11047;			
DT	01-JUL-1989 (Rel. 11, created)			
DT	01-NOV-1991 (Rel. 20, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Laminin gamma-1 chain precursor (Laminin B2 chain).			
GN	LAMC1 OR LAMB2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.			
OX	NCBL-TAXID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=8819845; Pubmed=1985895;			
RA	Pikkariainen T., Kallunki T., Ikonen J., Chow L.T., Tryggvason K.; "Structure of the human laminin B2 chain gene"; divergence from the laminin B1 chain gene.";			
RT	J. Biol. Chem. 266:221-228(1991).			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=91093128; Pubmed=1985895;			
RA	Kallunki T., Ikonen J., Chow L.T., Tryggvason K.; "Structure of the human laminin B2 chain gene"; divergence from the laminin B1 chain gene.";			
RT	J. Biol. Chem. 263:6751-6758(1988).			
RL	[3]			
RP	SEQUENCE OF 1393-1609 FROM N.A.			
RX	MEDLINE=89156633; Pubmed=3234037;			
RA	Fukushima Y., Pikkariainen T., Kallunki T., Shows T.B.; "Isolation of a human laminin B2 (LAMB2) cDNA clone and assignment of the gene to chromosome region 1q25->q31.";			
RA	Haley L.L., Henry W.M., Tryggvason K.; "Assignment of the gene to chromosome region 1q25->q31.";			
RT	Cytogenet. Cell Genet. 48:137-141(1988).			
RL	[4]			
RP	SEQUENCE OF 1382-1609 FROM N.A.			
RC	TISSUE-Endothelial cells;			
RX	MEDLINE=9221629; Pubmed=1806043;			
RA	Santos C.L.S., Sabbaga J., Brentani R.; "Differences in human laminin B2 sequences.";			
RT	"Differences in human laminin B2 sequences.";			
RL	DNA Seq. 1:275-277(1991).			
CC	-I- FUNCTION: Binding to cells via a high affinity receptor, laminin is thought to mediate the attachment, migration, and organization of cells into tissues during embryonic development by interdigitating with other extracellular matrix components.			
CC	-I- SUBUNIT: Laminin is a complex glycoprotein, consisting of three different polypeptide chains (alpha, beta, gamma), which are bound to each other by disulfide bonds into a cross-shaped molecule comprising one long and three short arms with globules at each end.			
CC	THE GAMMA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ), LAMININ-2 (MEROSENIN), LAMININ-3 (S-LAMININ) AND LAMININ-7 (KS-LAMININ).			
CC	LAMININ-6 (K-LAMININ) AND LAMININ-7 (KS-LAMININ).			
CC	-!- SUBCELLULAR LOCATION: Extracellular.			

ALIGMENTS

Result No.	Query	Match Length	DB ID	Description
1	251	4.6	I_MG1_HUMAN	P11047 homo sapien
2	225	4.1	I_MG1_MOUSE	P02468 mus musculus
3	224.5	4.1	I_MG1_DROME	P21849 giardia lamblia
4	211.5	3.8	I_TS4_GIAIA	P91066 mus musculus
5	209.5	3.8	I_LMG3_MOUSE	Q95922 mus musculus
6	209	3.8	PCK5_BRACL	Q9ijl2 giardia lamblia
7	208.5	1696	I_TS1_GIAIA	VS41_GIAIA
8	193	3.5	I_TS1_GIAEL	Q18823 caenorhabditis elegans
9	192	3.5	I_LMG3_HUMAN	Q9Y666 homo sapiens
10	192	3.5	I_LMB1_HUMAN	P07842 homo sapien
11	174	3.2	I_LMA2_MOUSE	Q60675 mus musculus
12	172.5	3.1	I_TS1_GIAIA	Q9ijl2 giardia lamblia
13	171.5	3.1	I_LMB1_MOUSE	P02469 mus musculus
14	170	1786	I_FUR2_DROME	P310432 drosophila melanogaster
15	169	3.1	I_TS1_GIAEL	P06882 rattus norvegicus
16	168.5	3.1	I_TS1_GIAIA	P16581 homo sapien
17	167	3.1	I_LEM2_HUMAN	P24043 homo sapien
18	166.5	3.0	I_LMA1_MOUSE	P19137 mus musculus
19	163	3.0	I_TS044_MOUSE	P34504 caenorhabditis elegans
20	162	2.9	I_YMV2_GIEEL	P21783 xenopus laevis
21	161.5	2.9	I_NTC2_XNLIA	P11046 drosophila melanogaster
22	160.5	2.9	I_LMB1_DROME	Q21313 caenorhabditis elegans
23	160	2.9	I_LM2_CEEEL	P54754 mus musculus
24	160	2.9	I_NTC2_MOUSE	Q35156 mus musculus
25	159	2.9	I_YKB2 YEAST	P36170 saccharomyces cerevisiae
26	159	2.9	I_JAG1_HUMAN	P77804 homo sapien
27	158	2.9	I_FBN1_BOVIN	P98133 bos taurus
28	158	2.9	I_TS1_DROME	Q00174 drosophila melanogaster
29	158	2.9	I_LMA1_HUMAN	Q13753 homo sapien
30	157.5	2.9	I_FBN2_MOUSE	Q61556 mus musculus
31	157.5	2.9	I_FBN2_HUMAN	P35556 homo sapien
32	157.5	2.9	I_LMA1_HUMAN	P25391 homo sapien

Query Match	4.6%	Score 251; DB 1; Length 1609;	Best Local Similarity	19.6%; Pred. No. 4.1e-10;
Matches	212;	Conservative	85;	Mismatches 382; Indels 400; Gaps 54;
Qy	23	PRLWLLILAGTARQVT-----QGTGELHACKS-----EYKVEYTAGDSCSRW 68	RESULT 2	
Db	130	PSINLTLHIGKANDITVYRKFHTSRSEFAVKRTREDGPPIPQYQYSGCENTVSKA 189	LMGL_MOUSE	
Qy	69	RVAVPHTPG----LCTSLPDYKGTECSFSCNAGEFLIDMKDOSCKPCAEGRYSLGNGIR 123	ID_LMGL_MOUSE	STANDARD;
Db	190	NRGFRTRIGGDEQALCDB-----EFSDF---SPFPGWAFSI-- 224	PRT;	1607 AA.
Qy	124	FDEMDELPHGFAISLSANMFLDDSAESTGNCTTSKWKWVGRGYIAFNDECTATLMLYAVNL 183	PO2458;	
Db	225	-----LEGRPSAYNFNDSPVQEWVATD----- 253	DT	21-JUL-1986 (Rel. 01, Created)
Qy	184	KQSGTVNFE-----YVDSLIEFFVONDOPNADSRMKTTERKGWEHSV 233	DT	01-JUL-1989 (Rel. 11, Last sequence update)
Db	254	NRLWHPFGDEVENDPKVLKSYYYAISD----FAVGGRKCNGHASECMKN----EFDKL 303	DE	15-JUN-2002 (Rel. 41, Last annotation update)
Qy	234	ELNRQNNVL-----YWRTAFTSWTKVPKVLVRNIAITGVAYTSECPVC 278	GN	Laminin gamma-1 chain precursor (Laminin B2 chain).
Db	304	VNCNHKHTVQVDCEKCLPFNDPWRRTAES----- 342	OS	LAMC1 OR LAMC1 OR LAMB-2.
Qy	279	-----KPGH-----ADKQSSFKCLCPANSY--SNKGGETSCHOCD 312	OC	Mus musculus (Mouse).
Db	343	DCGRSQECYFDPPELYRSTGHGHCNTNCODNTGACERCERCRENFRGLNNEACSSCH-CS 401	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Qy	313	P-----DKYSERKGSSCNVRA-----CTDKYFYT-----ACDNGEholm 351	OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Db	402	PVGSLSTQCDSYGR----CSCKPGVGMGDKCDRCQGPFGHSLTAGCRPCSCDCDGSSTDEC 456	NCBI_TaxID=10090;	
Qy	352	YKKWAKPKICSDELGAHKVLPASGVKTHCPCPNCOFFKTNST--CQCPYGSSNSDC 408	RN	[1]
Db	457	NVEIGRCVCKDNFEG-----ENFCRCKGFNFENNSNRRGCTC--RCFGHSSVC 504	RN	SEQUENCE FROM N.A.
Qy	409	TRCPAGTEPAVGFFEYKWNNTLPTNMETTVLGLNINFEXYKMTGWEVAGDHITYAGASND 468	RX	MEDLINE=88059118; PubMed=3680290;
Db	505	TN-----AVGYSV-----YSISSTFOIDEDGWRAJORDGEASLENSSERQD 546	RX	Sasaki M., Yamada Y.; "the laminin B2 chain has a multidomain structure homologous to the
Qy	469	FMLT-----LVPGERPQSYMADNTKEARIVTFET--LCVNCYELYFMGVN 518	RX	"the laminin B2 chain promoter contains unique repeat sequences and
Db	547	TAIVSDSFYPRYTAFAKFLGKVLSYGN----LSFSFRVRDRDTLSEADEVLEGAG 601	RX	is active in transient transfection.";
Qy	519	SRTNPVTPWKGSKGKQSYTIITENTTSFTW-----AQ-----RWFH 559	RX	Barlow D.P., Green N.M., Kurkinen M., Hogan B.L.M.; "the laminin B2 chain promoter contains unique repeat sequences and
Db	602	LRVSPVPLAGNSUPSETTVKYVHLREATDYPURPALPPEFOQLLNLTISKIRGYS 661	RX	is active in transient transfection.";
Qy	560	EASKYNTDYAKT----- 576	RX	Sequencing of laminin B chain cDNAs reveals C-terminal regions of
Db	662	ERSAGYLDDTLASARPGPQVATWESCTCPVGSYGGQFCMISLGSYRETPNLUGPSPC 721	RX	RT
Qy	577	VINVANGVASYCRP-----ALEASDVGSCSTCPCAGYYIDRD-SGT--CHSCC-PNTLIR 628	RX	coiled-coil alpha-helix.";
Db	722	VLCACNGHSETCDPTEGVNCRDNTAGPHCEKCSIDGYYGDSTACTSSDCQPCPGSSC 781	RX	EMBO J. 3:2355-2362(1984).
Qy	629	AHQPYQVQ-ACYPCPGPGTKNNKIHSLC--YNQDFTSRMRPT-----TFWVNSALA 677	RX	[5]
Db	725	EGESGSKSTIAYVCAVITIPPEVYKAGV-----SSQPSLAD-RLTGVTDM- 773	RX	X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 771-932.
Qy	782	AVWPKTEKVYCTNCPTGTCRKC-ELCDCGYYGPGLGRGPRVCRLLCOCSDNDPNAVG 840	RX	MEDLINE=96196434; PubMed=8648630;
Db	678	NTVTLAGGPSFTSKGKYFHH-----FTLSCLGNNQRKMSVCTDNTDLRIP 724	RX	Steinfeld J., Mayer U., Timpl R., Huber R.;
Qy	841	NCNRLTG----ECUKCINTAGFYCDRCKDGFGNPLAPNAPDKCKACNCN-----P 888	RX	RT
Qy	889	YCTMKOOSCNPVTGOCECL-PHWIGQDCGACDFCFYINLQSOGCCRDCHALGSTNGOC 947	RX	"Crystal structure of three consecutive laminin-type epidermal growth factor-like (LE) modules of laminin gamma chain harboring the nidogen binding site."
Db	774	-----TLDGITSPAELHLESGLPDYFIFRSRNDTVOSSSGRSSTIRVRCSPKT 825	RX	RT
Qy	948	DIRTGOECOPCGTGO---HCERC---EVNHFGRGPESCKPCD-----CHPE-- 988	RX	J. Mol. Biol. 257:658-668(1996).
Db			STRUCTURE BY NMR OF 824-881.	
Qy			MEDLINE=96196435; PubMed=8648631;	
Db			Baumgartner R., Czisch M., Mayer U., Poeschl E., Huber R.,	
Qy			Timpl R., Holak T.A., "Structure of the nidogen binding LE module of the laminin gamma-1 chain in solution.";	
Db			J. Mol. Biol. 257:658-668(1996).	
Qy			- FUNCTION: Binding to cells via a high affinity receptor, laminin is thought to mediate the attachment, migration, and organization of cells into tissues during embryonic development by interacting with other extracellular matrix components.	
Db			CC - SUBUNIT: laminin is a complex glycoprotein, consisting of three different polypeptide chains (alpha, beta, gamma), which are bound to each other by disulfide bonds into a cross shaped molecule	
Qy			Comprising one long and three short arms with globules at each end	
Db			THE GAMMA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ), LAMININ-3 (S-MEROSIN), LAMININ-4 (S-MEROSIN),	

LAMININ-6 (K-LAMININ) AND LAMININ-7 (KS-LAMININ).
 -! SUBCELLULAR LOCATION: EXTRACELLULAR.
 -! TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
 COMPONENT).
 -! DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
 WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
 -! DOMAIN: DOMAINS VI AND IV ARE GLOBULAR.
 -! SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
 -! SIMILARITY: CONTAINS 11 LAMININ EGF-LIKE DOMAINS.
 -! SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

FT DISUFID 443 454 BY SIMILARITY.
 FT DISUFID 445 461 BY SIMILARITY.
 FT DISUFID 463 472 BY SIMILARITY.
 FT DISUFID 475 490 BY SIMILARITY.
 FT DISUFID 722 731 BY SIMILARITY.
 FT DISUFID 724 738 BY SIMILARITY.
 FT DISUFID 740 749 BY SIMILARITY.
 FT DISUFID 752 768 BY SIMILARITY.
 FT DISUFID 771 779 BY SIMILARITY.
 FT DISUFID 773 790 BY SIMILARITY.
 FT DISUFID 793 802 BY SIMILARITY.
 FT DISUFID 805 823 BY SIMILARITY.
 FT DISUFID 826 840 BY SIMILARITY.
 FT DISUFID 828 847 BY SIMILARITY.
 FT DISUFID 850 859 BY SIMILARITY.
 FT DISUFID 862 879 BY SIMILARITY.
 FT DISUFID 882 896 BY SIMILARITY.
 FT DISUFID 884 903 BY SIMILARITY.
 FT DISUFID 905 914 BY SIMILARITY.
 FT DISUFID 917 930 BY SIMILARITY.
 FT DISUFID 932 945 BY SIMILARITY.
 FT DISUFID 935 952 BY SIMILARITY.
 FT DISUFID 954 963 BY SIMILARITY.
 FT DISUFID 966 978 BY SIMILARITY.
 FT DISUFID 981 993 BY SIMILARITY.
 FT DISUFID 983 999 BY SIMILARITY.
 FT DISUFID 1001 1010 BY SIMILARITY.
 FT DISUFID 1013 1026 BY SIMILARITY.
 FT DISUFID 1029 1029 BY SIMILARITY.
 FT DISUFID 1032 1032 INTERCHAIN (PROBABLE).
 FT DISUFID 1598 1598 INTERCHAIN (WITH CHAIN BETA-1).
 FT CARBOHYD 58 58 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 132 132 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 574 574 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 648 648 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1020 1020 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1105 1105 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1159 1159 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1173 1173 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1203 1203 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1221 1221 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1239 1239 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1378 1378 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1393 1393 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1437 1437 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 216 216 G -> A (IN REF. 3).
 FT CONFLICT 260 260 E -> D (IN REF. 2).
 FT CONFLICT 337 337 S -> C (IN REF. 2).
 FT CONFLICT 447 448 LR -> PS (IN REF. 2).
 FT CONFLICT 544 544 D -> Y (IN REF. 2).
 FT CONFLICT 662 662 T -> S (IN REF. 2).
 FT CONFLICT 886 886 MISSING (IN REF. 2).
 FT CONFLICT 1158 1158 MISSING (IN REF. 2).
 FT CONFLICT 1434 1434 V -> A (IN REF. 2).
 FT CONFLICT 1475 1475 R -> K (IN REF. 4).
 FT CONFLICT 1576 1576 D -> N (IN REF. 4).
 SQ SEQUENCE 1607 AA; 177297 MW; 81B7B08E4869F242 CRC64;

Query Match 4.1%; Score 225; DB 1; Length 1607;
 Best Local Similarity 19.6%; Pred. No. 3.1e-08;
 Matches 212; Conservative 83; Mismatches 368; Indels 416; Gaps 53;

QY	23	PRWLRLIIGAFAQVT-----QGTGPELLHACKES-----EYHEYEYACDSTGSRW	68
Db	128	PNSINLTHLGRAFDITVVLKFHTSRPESFAIYKRREDGPMWIPQYSSCVENTYSA	187
QY	69	RVAVPHTG-----LCVSLPDVPGKTCSCSFNAGERFMDKPSCKCAEGVSLGIR	123
Db	188	NRGFIRTGGDGDEQALCTD-----EFSDI-----SPLTGKNAFST-----	222
QY	124	FDEWDELPHGAFASLANSNELDSAESTGNCTSSKWPGRGYIAFNDECATLAVNL	183
Db	223	LEGRPSAYNFNDNSPVLOQWYTATD-----IRVTL	251

Wed Mar 12 10:08:37 2003

RC STRAIN=AD-1;
 RX MEDLINE=9334970; PubMed=8325510;
 QY 306 TSCHQCPDPKYSEKGSSCNVRPACTDKDYFTHTACDANGETQLMYKWAKPKICSEDE 365
 Db 456 HGGQCGCD----SGGSQNT-PAC----- 489
 QY 366 GAVKLIPASGVKTCPPCNPGFF---KTNNSTCQPCPYGSNSDCOTRCAGTEPAV--- 419
 Db 490 G-----RCNECKPGFFFLNDKNNRFGCTPC-FCYGHTECMATGCVSUTSN 537
 QY 420 --GFFEYKWNTLPIPMETMVLSCINFEYKGMCWEGAG-DHYTAAGASDNDFMLTVV 476
 Db 538 FNKKERWTAADLNOREVDIKY-N-QYSRIGTTAOGNERHV----- 577
 QY 477 PGFRPPQSWAD--TENKEVARITFVETLCSCNCYLFWMGVNSRTNPVETWGS-- 531
 Db 578 --FQARDFLGDRASYNDRDLKFLQLYQVALNGVSVDIVLEGAGSRISLPIFA-QGNGI 634
 QY 582 NGV-----ASYCRPCALEASDVGSCTSCPAGYIDRDSG-----TCH 619
 Db 688 DVELQTAHRGAGHAFPATWIEQTCIPEGVYQGOTCEASCAPGYRHSARGPFEMCIPCDH 747
 QY 620 S---CIPNT--ILKAHOPYGWQACYCVPGPGTKNNKIHSLCYNDCTFSRNTPRTFNYNF 673
 Db 748 GHADICDSETGRGCICOHNTG-DNCDOCAKGFYGNALGG-TPNPC--KRCP----- 794
 QY 674 SALANTVLAGPSFTSKGLKRYFHHTLSCGNGRKMSVCDTNDTDLRPG----- 726
 Db 795 -----C-PNDGACLOQNEDIVTCBPKSYFGSRC 824
 QY 727 -ESGFSKSITAVY---CQAV---IIPPEVTGKAGYSSQPSLADRLLGVTDMTL 775
 Db 825 QCSQGFGFDPTGIGLGEVOTCKSCDCNGNDPNAVGNCNRGTRGECLKCIIRNTAGHCQCL 884
 QY 776 DG-TSPAEELFHIESLGLGPDVIFYRSNDVQSCSSGRSTIRVRCSPORTVPGSLLPG 834
 Db 885 SGHFGDPLPH---GRCDRSCYEAQTEQEPEQSITRCQDTGOCQKPNVIGRDC-G 938
 QY 835 TCSDG-----TCDGCONFHEWESA 853
 Db 939 ECQPGYFNTRSGNGCENCILCDPVGSYNSTCDRYSGQCHCRPGVMGORDOCENYFYGESS 998
 QY 854 AACPIC 859
 Db 999 BGCKPC 1004
 RESULT 4
 TSEA_G1ALA G1ALA STANDARD; PRT; 713 AA.
 ID P21849; DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-MAY-1991 (Rel. 18, last sequence update)
 DT 16-OCT-2001 (Rel. 40, last annotation update)
 DE Major surface-labelled trophozoite antigen 417 precursor.
 GN TSA 417.
 OS Giardia lamblia (Giardia intestinalis).
 OC Eukaryota; Diplomonadida; Hexamitidae; Giardinae; Giardia.
 OX NCBI_TAXID=5741;
 RP SEQUENCE FROM N_A.
 [1] SEQUENCE FROM N_A.
 RC STRAIN=ATCC 30957 / WB;
 RX MEDLINE=928939; Published=2352929;
 RA Gillin F.D., Haggblom P., Harwood J.J., Aley S.B., Reiner D.S., McCaffery M., So M., Guiney D.G.;
 RT "Isolation and expression of the gene for a major surface protein of
 RL Giardia lamblia.";
 RT Proc. Natl. Acad. Sci. U.S.A. 87:4463-4467(1990).
 RN [2] SEQUENCE OF 480-620 FROM N_A.
 RP

Db 373 ---- SCAKIGNYGATEGAK-KLCKE----CTAANC KTCDDQGCQACNDGFYK- 417
 QY 390 NNSTCOPCPYG---SYNSGSDCIRCPAGTEPAPVGFYKWNTLPTNMETTVLSGTFNEY 445
 Db 418 NGDACSPCHESCKTCSAGTASDCTECPTG-----KALRIGD 453
 QY 446 KGMTGNGEVAGHIVTAGASDENDFMLTLWPGF-----RPPQSMADTEN--EV 494
 Db 454 DGTKG-TCGGCTTGAGACKTGTGCLTDGASYCSECATTEPYQONGVCPKASRATPT 511
 QY 495 ARITVFETICSVNCLELYN- -GVNSRNTNPVETW----KGSKKOSYVILENTT 546
 Db 512 CNDSPIONQVGCTCADNYFKMMNGCYEVSKYRKTVCISAPNGTCOKAAGDYKLDSGL 571
 QY 547 TSFTWAQORTTFHEASRKVTNDYAKIYINVNTVMGNY--ASYCRPCALE--ASDVG 600
 Db 572 -----TVCSEGCKECAST----STDCITCLDGVKSACTKCDASCETCNGAA 616
 QY 601 SCTSOPAGY-IDRSGRCHSCCPNTILKAHPGYQACVPGCPGPKNNKTHSLCY 655
 Db 617 TCKACATGYKKTASGEACTS ESD---SNGVTGIRGCLNCAP-PPNNKGSVLCY 667

RESULT 5
 LMG3_MOUSE STANDARD:
 ID LMG3_MOUSE PRT; 1581 AA.
 AC Q9RQB6; O9WWW6;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE Laminin gamma-3 chain precursor (Laminin 12 gamma 3).
 GN LAMC3.
 OS MUS musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Albus A.M., Burgesson B., Champaliard M.-F., Koch M., Olson P.;
 RT "Mouse laminin 12 gamma 3 chain.";
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 RR [2]

SEQUENCE OF 1-1526 FROM N.A.
 RX MEDLINE=9253369; PubMed=10318827;
 RA Iivanainen A., Morita T., Ryggvason K.;
 RT "Molecular cloning and tissue-specific expression of a novel murine laminin gamma 3 chain.";
 RL J. Biol. Chem. 274:14107-14111(1999).

-!- FUNCTION: Binding to cells via a high affinity receptor, laminin is thought to mediate the attachment, migration, and organization of cells into tissues during embryonic development by interacting with other extracellular matrix components.

-!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three different polypeptide chains (alpha, beta, gamma), which are bound to each other by disulfide bonds into a cross-shaped molecule comprising one long and three short arms with globules at each end.

THE GAMMA-3 CHAIN IS A SUBUNIT OF LAMININ-12.

-!- SUBCELLULAR LOCATION: Extracellular.

-!- TISSUE SPECIFICITY: Strongly expressed in capillaries and arterioles of kidney as well as in interstitial Leydig cells of testis.

-!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.

-!- DOMAIN: DOMAIN IV IS GLOBULAR.

-!- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).

-!- SIMILARITY: CONTAINS 11 LAMININ EGF-LIKE DOMAINS.

-!- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/>) or send an email to license@isb-sib.ch).

CC modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>) or send an email to license@isb-sib.ch).

CC or send an email to license@isb-sib.ch).

EMBL: AF033372; AAC08983.1; -.

DR EMBL; AF03520; AAC29851.1; -.

DR HSSP; P02468; 1KLO.

DR MGI; MGI-1344394; Lamc3

DR InterPro; IPR00561; EGF-like.

DR InterPro; IPR00186; Lamint.

DR InterPro; IPR00034; Laminin_B.

DR InteraPro; IPR002049; Laminin_EGF.

DR Pfam; PF00052; Laminin_B; 1.

DR Pfam; PF00055; Laminin_EGF; 10.

DR PRINTS; PRO001; EGFLAMININ.

DR ProDom; PD00282; Lamnt; 1.

DR SMART; SM00180; EGF_Lam; 10.

DR SMART; SM00281; LamB; 1.

DR SMART; SM00136; LamNT; 1.

DR PROSITE; PS00022; EGF_1; UNKNOWN_8.

DR PROSITE; PS01186; EGF_2; 2.

KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil; Laminin EGF-like domain; Cell adhesion; Repeat; Signal.

FT SIGNAL 1 28 POTENTIAL.

FT CHAIN 29 1581 LAMININ GAMMA-3 CHAIN.

FT DOMAIN 29 1581 LAMININ N-TERMINAL (DOMAIN VI).

FT DOMAIN 280 335 LAMININ EGF-LIKE 1.

FT DOMAIN 336 391 LAMININ EGF-LIKE 2.

FT DOMAIN 439 438 LAMININ EGF-LIKE 3.

FT DOMAIN 439 488 LAMININ EGF-LIKE 4.

FT DOMAIN 489 498 LAMININ EGF-LIKE 5 (N-TERMINAL).

FT DOMAIN 685 718 LAMININ DOMAIN IV LAMININ EGF-LIKE 5 (C-TERMINAL).

FT DOMAIN 719 766 LAMININ EGF-LIKE 6.

FT DOMAIN 767 821 LAMININ EGF-LIKE 7.

FT DOMAIN 822 877 LAMININ EGF-LIKE 8.

FT DOMAIN 878 927 LAMININ EGF-LIKE 9.

FT DOMAIN 928 975 LAMININ EGF-LIKE 10.

FT DOMAIN 976 1024 LAMININ EGF-LIKE 11.

FT DOMAIN 1025 1581 DOMAIN II AND I.

FT DOMAIN 1029 1046 COILED COIL (POTENTIAL).

FT DOMAIN 1112 1153 COILED COIL (POTENTIAL).

FT DOMAIN 1210 1231 COILED COIL (POTENTIAL).

FT DOMAIN 1438 1458 COILED COIL (POTENTIAL).

FT DOMAIN 1510 1575 COILED COIL (POTENTIAL).

FT CARBHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBHYD 304 304 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBHYD 337 337 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBHYD 640 640 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBHYD 849 849 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBHYD 991 991 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBHYD 1162 1162 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBHYD 1196 1196 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBHYD 1320 1320 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBHYD 1514 1514 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CONFLICT 9 9 L-> F (IN REF. 2). P-> T (IN REF. 2). R-> K (IN REF. 2).

FT CONFLICT 190 190 R-> K (IN REF. 2).

FT CONFLICT 195 195 G-> S (IN REF. 2). C-> R (IN REF. 2). C-> Y (IN REF. 2).

FT CONFLICT 221 221 P-> T (IN REF. 2). R-> K (IN REF. 2).

FT CONFLICT 394 394 R-> K (IN REF. 2).

FT CONFLICT 471 471 C-> Y (IN REF. 2).

FT CONFLICT 1150 1150 L-> DDEPOLSLILK (IN REF. 2).

FT CONFLICT 1387 1387 H-> Q (IN REF. 2).

FT CONFLICT 1438 1439 AS-> T (IN REF. 2).

FT CONFLICT 1479 1479 V-> I (IN REF. 2).

SEQ 1581 AA; 172316 MW; 51DAD1F95E6AE81 CRC64;

Query Match Best Local Similarity Score Length
 Matches 218; Conservative 107; Mismatches 394; Indels 451; Gaps 60;

Wed Mar 12 10:08:37 2003

QY	23 PRLLRLLWAGIAFOVT-----OGTGPMLACKE-----SEHYEYTACDST--- 64
Db	124 PTSYNTLISIGKAYEITSVRLKFHTSRPESIAYKRTVASYGWPWEPYQYASASCOTYGRP 183
QY	65 -GSRWVAVFHTRLCIT-LPD----PVGTEESFSCHNAGEFIELDMKDOSCKPCAEVSLGT 120
Db	184 EGHYLRPGEDERVFATCTSFSDIPLSPLNGNWAES-----TLEGRPS--- 224
QY	121 GIRFDEWDELPHGFASLSAMELDDSAESTGNCTSSKWVPRGYIAFNDECATIMYA 180
Db	225 AYNFEESPVL----- 254
QY	181 VNLKQSGTIVNFYVYDPDSIIFEFV-----ONDOPCPNA----- 215
QY	255 DDIFKDPRVILQSYY---AVSDFSVGGRKCKNGHASECEPNAGQLACROHNTGVD 310
Db	356 -----GHHGCORCR---DHITGPIGERCEBKNNYRWSPKTPCOPCD---CHPAG 398
QY	216 -----DDSRMKTTEK-----GWEETHSELNRNNNLYWRTHAFSWMTKVKPKPV 355
Db	311 ERCLPFFODRPWARGTAEDANECLPCNCSCSHSECFTDREL---YRST----- 320
QY	261 VRNIALTGIVAYUSECFCPKPGTYADKGQSSFKLCPANSYNSKGETSCHOCDPKYSEKG 320
Db	399 SLSLGQ-----CDNSG-----VC-----PCKPTVIGK---CD 422
QY	381 PCNPGEFKVNNTICQ-----CPVGSVNSGSDCTRCACT----- 415
Db	423 RCLPGFHSLSEGGCRPCACNVAGSLGTCPRSGNCPCKEYNEVGSLCDCRCRPGTFENQPHN 462
QY	416 -----EPVAGFEKWWNHLPTNMETTVLGLINFEYKGMOWEVAGH----- 458
Db	529 KRLQWSQSGLGLRGGEELSAPKKFLGDLRSLYQGPWLITQVPGGSPPIQLREG 588
QY	483 -----QSWHAD-ENKEVARITVEETICSVNCELYFMV 515
Db	589 AGLAISLPRSSLSPQDTRPRVQLQFLQETSEAEASPLPTHQRLS----- 639
QY	516 GVSNSRTNTPTVETWKGSKGKOSYTIIEENTTSFWAFORTHEAS-----RKYTN 567
Db	640 -----NLTAISIWTSQGQVQCLLCEVOLTSAMP-QRLBLAPPASWVPC 693
QY	568 DVAKATSYINTVNM---NSVASYCORPCLAEASV----- 753
Db	694 QFCERCALGKREPRPHGJYANGIIPCNCQHETCQDPNMGICLCLGHTEGFSCKERMPGY 753
QY	611 IDRDSG---"ICHSCP-PNTILKAHQP-YGVQACVCPGCGFTKNNKHSCLNDCTSFRNTP 665
Db	754 GRAFSGRADQOPCPCPGOSACATIPESGDMVYTHCPGQRRCES-CED----- 803
QY	722 RIEGESEFSKSITATWCQAVITIPPEVTGKAGYSSQPV-----LARLIGTTDMILDG 777
Db	854 HCEHCRGEG----- 893
QY	778 ITSPAEFLHESLGIPIY-----IFFYRSNDVTOSCSSGRSTIRVYCSPOKTV 826
Db	804 -----GFFGDPLGLSGAQ-PCCRQCSGNVLDNAVGNCDPHGSHCLR-CLYNTIGA 853
QY	894 VTVG-----OCVCLPYYSGRGDRDCSRCSPGFDLQSGRCQSKCHPLGSLENKCHPK--- 943
Db	827 PGSLLPLGTC-----SPGTCDCGCFNHFILWESAAACPLCSVADYHAIVSSCVAGIQRTV 880
QY	944 -----TGCCPCRPVGTVGOACDRQCLGGFFGSIKGCFDCRSPLAGASSOC-- HENRIC 994
Db	881 WREPPKICSGGSLISPERQRTICKIDFWMLKVG-ISACTCTATLTLVTCFWKKNQKLEY 939
QY	995 VCR-----PGFGVYKCDR---COD-NFLAGDTGQOEC-----TCVALYKEAKL 1050
Db	940 KYSLVMMATLK-----DCDUEAADSCAIMEGE 967
QY	1039 KARLMALMEWMQSLRSDCGSPWG-PLDILQGE 1067
RESULT 6	
PC5_MOUSE	STANDARD; PRT: 1877 AA.
ID	PC5_MOUSE
AC	Q04592; 062040;
DT	01-FEB-1995 (Rel. 31, Created)
DT	16-OCT-2001 (Rel. 40, Last sequence update)
DT	15-JUN-2002 (Rel. 41, Last annotation update)
DE	Proprotein convertase subtilisin/kexin type 5 precursor (EC 3.4.21.-)
DE	(Proprotein convertase PC5) (Subtilisin-like proprotein convertase 6)
DE	(convertase PC5) (Subtilisin-like proprotein convertase (PC6)).
DE	PC5.
DN	Mus musculus (Mouse).
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX	NCBI_TaxId=10090;
RN	[1] SEQUENCE OF 330-1877 FROM N.A. (ISOFORM PC5B).
RP	SEQUENCE FROM 330-1877 FROM N.A. (ISOFORM PC5A).
RC	STRAIN-ICR; TISSUE-intestine;
RC	LINEAGE-93227934; PubMed-8351106;
RA	Nakagawa T., Murakami K.;
RA	"Identification of an isoform with an extremely large Cys-rich region of PC6, a Kex2-like processing endoprotease.";
RT	FIBS Lett. 327:165-171(1993).
RL	[2]
RN	SEQUENCE FROM N.A. (ISOFORM PC5A).
RP	SEQUENCE FROM N.A. (ISOFORM PC5A).
RC	TISSUE-Brain, and Intestine;
RC	LINEAGE-93224489; PubMed-8468318;
RA	Nakagawa T., Hosaka M., Torii S., Watanabe T., Murakami K.,
RA	"Identification and functional expression of a new member of the mammalian Kex2-like protease family: its striking structural similarity to PACE4";
RT	J. Biochem. 113:132-135(1993).
RL	[3]
RN	SEQUENCE FROM N.A. (ISOFORM PC5A).
RP	SEQUENCE FROM N.A. (ISOFORM PC5A).
RC	TISSUE-Adrenal cortex;
RC	LINEAGE-9324056; PubMed-8341687;
RA	Ljsson J., Vieau D., Hamelin J., Day R., Chretien M., Seidah N.G.;
RT	"cDNA structure of the mouse and rat subtilisin/kexin-like PC5: a candidate proprotein convertase expressed in endocrine and nonendocrine cells.";
RT	Proc. Natl. Acad. Sci., U.S.A. 90:6691-6695(1993).
RL	[4]
RN	PARTIAL SEQUENCE, AND SUBCELLULAR LOCATION.
RP	LINEAGE-97103178; PubMed-8947550;
RX	De Bie I., Marciniuk Z.M., Malide D., Lazure C., Nakayama K.,
RA	Bendayan M., Seidah N.G.;
RA	"The isoforms of proprotein convertase PC5 are sorted to different subcellular compartments.";
RT	J. Cell Biol. 135:1261-1275(1996).
RN	[5]
RP	DEVELOPMENTAL EXPRESSION.
RA	LINEAGE-96293359; PubMed-8698813;
RT	Constam D.B., Calton M., Robertson B.J.;
RT	"SPC4, SPC5, and the novel protease SPC6 are coexpressed with bone morphogenetic proteins at distinct sites during embryogenesis.";
RT	J. Cell Biol. 134:181-191(1996).
RN	[6]
RP	DEVELOPMENTAL EXPRESSION.
RA	LINEAGE-97436919; PubMed-9291583;
RA	Rancourt S.L., Rancourt D.E.;
RT	"Murine subtilisin-like proteinase SPC6 is expressed during embryonic implantation, somitogenesis, and skeletal formation.";
RT	Dev. Genet. 21:75-81(1997).

CC -!- FUNCTION: LIKELY TO REPRESENT A WIDESPREAD ENDOPROTEASE ACTIVITY
CC WITHIN THE CONSTITUTIVE AND REGULATED SECRETORY PATHWAY. CAPABLE
CC OF CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF. MAY BE RESPONSIBLE
CC FOR THE MATURATION OF GASTROINTESTINAL PEPTIDES. MAY BE INVOLVED
CC IN THE CELLULAR PROLIFERATION OF ADRENAL CORTEX VIA THE ACTIVATION
CC OF GROWTH FACTORS.

CC -!- CATALYTIC ACTIVITY: RELEASE OF MATURE PROTEINS FROM THEIR
CC PROTEOPLAINS BY CLEAVAGE OF ARG-XAA-YAA-ARG- β -ZAA BONDS, WHERE XAA
CC CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS.

CC -!- SUBCELLULAR LOCATION: PC5A IS SECRETED THROUGH THE REGULATED
CC SECRETORY PATHWAY. PC5B IS EXPRESSED IN THE
CC A PARANUCLEAR POST-GOLGI NETWORK COMPARTMENT IN COMMUNICATION WITH
CC EARLY ENDOSOMES.

CC -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; PC5B/LONG (SHOWN HERE)

CC AND PC5A/SHORT; ARE PRODUCED BY ALTERNATIVE SPlicing.

CC -!- TISSUE SPECIFICITY: PC5A IS EXPRESSED IN MOST TISSUES BUT IS MOST
CC ABUNDANT IN THE INTESTINE AND ADRENALS. PC5B IS EXPRESSED IN THE
CC INTESTINE, ADRENALS AND LUNG BUT NOT IN THE BRAIN.

CC -!- DEVELOPMENTAL STAGE: WEAKLY EXPRESSED THROUGHOUT THE EMBRYO,
CC EXCEPT IN THE DEVELOPING NERVOUS SYSTEM, THE RIBS AND THE LIVER,
CC BUT MARKEDLY UPREGULATED AT DISCRETE SITES DURING DEVELOPMENT. AT
CC E7.5, PROMINENT EXPRESSION OBSERVED IN DIFFERENTIATED DECIDUA. AT
CC E7.5, INTENSE EXPRESSION IN EXTRAEMBRYONIC ENDODERM, AMNIOTIC AND
CC NASCENT MESODERM. AT E8.5, ABUNDANT EXPRESSION IN SOMITES AND YOLK
CC SAC FOLLOWED BY A CONFIRMATION TO DERMANYXON COMPARTMENT. BETWEEN
CC E9.5 AND E11.5, ABUNDANT EXPRESSION IN AER (THICKENED ECTODERMAL
CC CELLS OF LIMB BUDS). AT E12.5, EXPRESSION IN THE LIMBS IS CONFINED
CC TO THE CONDENATING MESOCHYM SURROUNDING THE CARTILAGE. AT THIS
CC STAGE, STRONG EXPRESSION ALSO DETECTED IN VERTEBRAL AND FACIAL
CC CARTILAGE PRIMORDIA AND IN THE MUSCLE OF THE TONGUE. AT E16.5,
CC ABUNDANT EXPRESSION IN THE EPITHELIAL CELLS OF THE INTESTINAL VILLI.
CC ISOFORM A IS MOST ABUNDANT AT ALL STAGES BUT SIGNIFICANT LEVELS OF
CC EXPRESSION ARE ALSO DETECTED IN THE MUSCLE OF THE TONGUE. AT E16.5,
CC ISOFORM B OCCUR AT E11.5.

CC -!- DOMAIN: THE PROPEPTIDE DOMAIN ACTS AS AN INTRAMOLECULAR CHAPERONE
CC ASSISTING THE FOLDING OF THE ZYMOGEN WITHIN THE ENDOPLASMIC
CC RETICULUM.

CC -!- DOMAIN: AC_1 AND AC_2 (CLUSTERS OF ACIDIC AMINO ACIDS) CONTAIN
CC WITH THE TGN SORTING PROTEIN PAC5-1.

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; D17583; BAA04507; 1; -;
DR D12619; BAA02143; 1; -;
DR EMBL; L14932; AAA04636; 1; -;
DR PIR; JX0248; JX0248;

DR PIR; A48225; A48225;

DR HSSP; Q98405; IMP1;

DR MEROPS; S08_076; -;
DR MGD; MG_97515; Pcsk5;

DR InterPro; IPR005651; EGF-like.

DR InterPro; IPR02174; Furin-like.

DR InterPro; IPR002884; P domain.

DR PIR; A48225; A48225;

DR Pfam; PF00082; Peptidase_S8; 1.

DR Pfam; PF01483; P; PARTIAL.

DR PRINTS; PR00723; SUBTILISIN.

DR InterPro; IPR002174; Furin-like.

DR SMART; SM00181; EGF; 3.

DR SMART; SM00001; EGF-Like; 2.

DR SMART; SM00261; FU; 22.

DR PROSITE; PS00136; SUBTILASE ASP; 1.

DR PROSITE; PS00137; SUBTILASE HIS; 1.

DR PROSITE; PS00138; SUBTILASE SER; 1.

KW Cleavage on pair of basic residues; Repeat; Alternative splicing;

KW Transmembrane; 1

FT SIGNAL 34

FT PROPEP 35 116

FT CHAIN 117 1877

FT DOMAIN 117 1768

FT DOMAIN 1759 1789

FT TRANSMEM 1790 1877

FT DOMAIN 117 452

FT DOMAIN 464 602

FT DOMAIN 638 1753

FT DOMAIN 1825 1844

FT DOMAIN 1855 1877

FT SITE 116 177

FT SITE 521 523

FT ACT_SITE 173 173

FT ACT_SITE 214 214

FT ACT_SITE 388 388

FT CARBOHYD 227 227

FT CARBOHYD 383 383

FT CARBOHYD 667 667

FT CARBOHYD 754 754

FT CARBOHYD 804 804

FT CARBOHYD 854 854

FT CARBOHYD 951 951

FT CARBOHYD 1016 1016

FT CARBOHYD 1220 1220

FT CARBOHYD 1317 1317

FT CARBOHYD 1523 1523

FT CARBOHYD 1711 1711

FT CARBOHYD 1733 1733

FT VARSPLIC 878 915

FT VARSPLIC 916 1877

FT SEQUENCE 1877 AA; 209287 MW; EC85022DF0EAC3 CRC64;

Query Match
Best Local Similarity 3.8%; Score 209; DB 1; Length 1877;
Matches 231; Consistency 90; Mismatches 399; Indels 446; Gaps 67;

QY 41 GRGPE-----LH-----ACKESEHYEVACDSTGSRHRVAVPHTPGLTS 81

Db 649 GPGPDHCSDCCHYYKKLNKNTRICVSSCPPGHYHADKRC-----RKAPNCECFG 701

QY 82 LPDPVKTCESFSCNAGELFDMDQSC-KPCVGEGRSIGLTGIRFDEWDDELPHGFLASLN 140

Db 702 -----HGDOC-LSKYGYFLNETSSCTQCDGSE-----DTKNNVKGKCSN 745

QY 141 MELDDSAAFSTGNCTSSKWW-----VPPGDIYAFNTDECATLMAVNLUQSQT-- 188

Db 746 CK----ACTGFHNOCTECKGGLSLSQGSRCSVICEDGFFNGHDQCPCHRFCATOSGAGDG 801

QY 189 -VMEIYVPPDSLIEFFVQNDQPNADSRWRMMITEKCGWEFHISVELNRGNVNWYRTT 247

Db 802 CINGTEGY-----VMEEGCVCOSVSYLDHSBG-GKSK-RCDNCLTNGP 850

QY 248 AFPSWTKVPKPVLVRNIAITGAWYVSECFCKPGTYADKGSSFKLCPAN----- 298

Db 851 GRINCSSCSPGIL---LGTCOMGAI--CKDGEYIDQG--HQQTEASACKCNGPTQ 902

QY 299 -----SYSNKGTSCHQCDPDKYERKGSSSCNVRPACTDK--- 333

Db 903 ECICISCPVTVRLLDDGRVMCNPSPNKFEPKKQCHPTCQSGQSGSOPNSNC'SCRADKHGQ 962

QY 334 DYFYTHAC-----DANGEL-----QLMWKWAKRKCSDLEGAVKIPASGYK 376

Db 963 ERFLYHGECLCENCPVGHYPAKGHNCLPCPDNGELCY--NPHICRCMSGVIVIPPNH- 1018

QY 377 THCP--PCNFGFKEFNN-STCQCPYGSYSNGSD---CITCPACTEPAVGFEKWNML 429

Db 1019 --COKLCRGEROFDSEYBECMCERCLGCTEDPGACSGCAG--YMFERHCYAC 1073

	Best Local Similarity	19.3%	Pred.	No.	5.3e-07;	
Matches	221;	Conservative	108;	Mismatches	399;	
Qy	49 CKSESEHYEYFACDSTGSRRVRAVPHFGCILSPDP--VKTECSFSCNAGEFIDMKDQ	415;	Indels	415;	Gaps	65;
Db	765 CSEGRMDENDVQCDDCDTCGPDATDCVCADEDLJLJESOCVESCSCGGYF--QQEY	106				
Qy	107 SCKPCAGEGRSLGTRGIRFEDDELPHGFASLSANMEELDSAESTGNCTSSKKWPRGYI	822				
Db	823 ECLKC-HATCASCSSGRDQ-----CLTGSGHLELD---EDTHRCITI-CEDGEY-	166				
Qy	167 AFNTDECTATLMFAVNQKOSGVNFEEVYPOSSIFFEYVONDQCPNADDENRMRKT-	867				
Db	868 --GTEEGKCEDCNICKNCNS-----	224				
Qy	225 -----EKWFHFHSV-----ELNRGN-----NVLIWRTAFSWTKV	907				
Db	908 QYCGNRRYIPENG-ECHPCHPSCLGCCIGGEINOCNQCIDYEGDHFENYQGT-----CHV	255				
Qy	256 PKFVLYVLENIAITGVAYTSECFCPKERGYA-DKOGSSFCKLK-----	950				
Db	961 TCPP-----GLYGDITDQVCKACARPGTACDGPADNOCTLCBEERAPTDGRGQSQT	302				
Qy	303 RGETS--CHOCDPDKYSEKGSSCNRVAPCTRKYFTHTAC-----	1015				
Db	1016 BAECFGCHSC-----FEGPFDIC--SC-DEDDYLTEDPCCRVRNNCPSTYDDQDR	347				
Qy	348 TOLMYKWARP-----KICSEDELAGVKLPGASVHKHP-----	1065				
Db	1066 CRPCHDNEACDPNNQNCNSKREGYKTP-DGCSTGCPNRYKDDTNKECKCDCSSCF	388				
Qy	389 TNIST--COPYPGISTSNGSC-TRCPACTEPAVGFEXKWWNLPTNNETWLSGINFE	444				
Db	1125 CSGPASFHCLSCADGFLPHESCCSRSTCAG-----	1124				
Qy	445 -----YKSMTGWEVAGHIIYAGASND-----FIMILTWPGFRPQSVMADETNKEVA	495				
Db	1171 QSYQYSETG--RCEDCPYNCRACDNNDGCACTCAPYIVDRCRRECEEDGEDYQDR	1227				
Qy	496 RITFVFETL-CSVNLCELYFMVGVMNSRTNPVFWKGSKGKOSYYTIEENTTSFTAFO	554				
Db	1228 RDTAELSCRPHQSKT----CSGPSPDIDCDSCKDD-----TILDRGRCITSC-----	1272				
Qy	555 RTTFHEASKYTVNDAVAKYSINTVNWVASY-----CRECALAEDSVGSSCT	603				
Db	1273 -----GPEBYMDRREKKKACHPTCKECSDEDDICT	1304				
Qy	604 SCPGYYIDRSDTCHS-CPPNLT-----KAHQ-----PYGVQACVPCFRGTRNK	649				
Db	1305 ACNDGFLL-TDASSCEAGCPGPFQFLHGDGDCSIRECKTCGPHH-DNCLSCQPS	1358				
Qy	650 IHSCLCYNCTFERRNTPTTFENFNESALANTVTLAAGGPSFTSKGLKVKFHFTLSSLCCNQGR	709				
Db	1359 ---YLNDQDSTCHCPEGTFBETYEDDGETVLIQ-----CRICHVNCKTERGEYE	1404				
Qy	710 KMSVCTDNVTLR-----IPEEBSGSKSTIAVCOAVIIPPEVYKGVSQPVSL	762				
Db	1405 E--DCMECANDIKYKQDGCVTEQEGHYPDLT-NECQQCWSDC	1446				
Qy	763 ADRLIGVTDMTLG-----TSPAEFLFHESLGLPDVYFYRSNDVTOCSSGRSTI	816				
Db	1447 -----TCGPGRNDQCVTCPPY-----YVLVAGKCLEDCPREGYDM	1482				
Qy	817 RV-----RSP-----KTP-GSLLPOTC-----	839				
Db	1483 RQEKEGCHPSCATCNGGNYNCSPYGSKLGPGVCYPMCEBEHEYVETOICERCND	1542				
Qy	840 TDCGNCNFLWLWSSAAAPLCSVY-DYHATVSSCVY--GIOKTY--WRREPKLC--	888				
Db	1543 SCKPCR--GSTAHOCILSCBAPYGHAMKHLCACCEBGSPENEYCCICHESTRICID	1598				
Qy	889 -----SGGISLSPSQRVFTICKIDFWLKVGISAGTGTAILTVL-----	926				
<hr/>						
Best Local Similarity						
3.5%; Score 193; Pred. No. 2.2e-06; Length 687;						
Db	1599 REAGVQFSSADSIPN-----VAYTAVATFICVVIVLFFVFGMLQARSNG	1646				
Qy	927 -TCYFWKKNNQKLEYKYSKLYNNATLKDCLDPADSCIMEGED-VREDDLFSTSKHSLR	984				
Db	1647 RLCWAHYQOQVPTTRYEK-MN-----DHVNILSQEDFYNEDS-SEDEIHSDS	1693				
Qy	985 SNH 987					
Db	1694 TRH 1696					
<hr/>						
RESULT 8						
VS41_GIALA						
ID	VS41_GIALA					
AC						
P9127;						
DT						
16-OCT-2001 (rel. 40, Created)						
DT						
16-OCT-2001 (rel. 40, Last sequence update)						
DE						
Variant-specific surface protein VS41AI precursor (CRISP-90).						
OS						
Eukaryota, Diplomonadida, Hexamitidae; Giardina intestinalis).						
OX						
NBL_TAXID=5741;						
RN						
SEQUENCE FROM N. A.						
RC						
STRATN=02-4A1;						
RX						
RA						
RP						
CHARACTERIZATION						
RA						
Papanastassiou P., Bruderer T., Li Y., Bommerli C., Koehler P.;						
"Primary structure and biochemical properties of a variant-specific						
surface protein of Giardia";						
RL						
Biomed. Parasitol. 86:13-27(1997).						
RX						
MEDLINE=91321554; PubMed=9178242;						
RA						
Papanastassiou P., McConville M.J., Ralton J., Koehler P.;						
"The variant-specific surface protein of Giardia, VS41AI, is a						
glycosylated and palmitoylated protein";						
RL						
Biochem. J. 322:49-56(1997).						
CC						
RT						
PLASMA MEMBRANE LOCATION: TYPE I MEMBRANE PROTEIN. ANCHORED TO THE						
CC						
-I- PTM: O-GLYCOSYLATED. THE MAJOR GLYCAN IS A TRISACCHARIDE WITH GLC						
CC						
-I- SIMILARITY: BELONGS TO THE GIARDIA VARIANT SURFACE PROTEIN FAMILY.						
CC						
This SWISS-PROT entry is copyright. It is produced through a collaboration						
between the Swiss Institute of Bioinformatics and the EMBL outstation -						
the European Bioinformatics Institute. There are no restrictions -						
use by non profit institutions as long as its content is in no way						
modified and this statement is not removed. Usage by and for commercial						
entities requires a license agreement. (See http://www.isb-sib.ch/announce/)						
CC						
DR						
DR_HSP; 203743; CAR06038; 1; -						
DR						
HSSP; 014763; 100G						
DR						
Glycosidase; P2127; -						
DR						
InterPro; IPR000561; EGF-like.						
DR						
InterPro; IPR021174; Furin-like.						
DR						
Pfam; PF03302; Giardia_VSP.						
DR						
SMART; SM00181; EGF; 2.						
DR						
SMART; SM0001; EGF-like; 1.						
DR						
SMART; SM0261; FU; 3.						
KW						
Antigen; Repeat; Transmembrane; Glycoprotein; Lipoprotein; Palmitate;						
FT						
SIGNAL	1	14				
FT						
CHAIN	15	687				
FT						
DOMAIN	15	660				
FT						
TRANSMEM	661	681				
FT						
DOMAIN	682	687				
SEQUENCE	687 AA;	8057 MW;	D892F675D62D7EC CRC64;			
Query Match	3.5%; Score 193; Pred. No. 2.2e-06; Length 687;					

us-10-046-433-40.rsp

Wed Mar 12 10:08:37 2003

		SEQUENCE FROM N.A.
Matches	149;	Conservative
		78; Mismatches
		243; Indels
		340; Gaps
		38;
QY	275	CFPKKPYTAQKGSSFCKLCPANSYS-----NKGEGTSCHOCDPDKY--- 316
Db	41	CTRCKTG---GVFPVDFGFCR---PPFGSIQAAAAGCTKADGTATCGKG---DGYFLFM 94
OY	317	-----SEKGSSSCNVR----PAC-TDKDYFVHTACDANGTOMLYKWRKPCKSED 363
QY	95	GCGYKTEQPGSIECTTASNGLCTACKVDSDYIFONKATPSEKSECILCW---TID 149
Db	364	LEGAVKLPLASGVKTHCPPCNPGFPTKNSTCOPCPVGSYNS---SDPCTRCPAGT 415
QY	150	RNGVM---GV-ANCATCTAPASSTGPATCPECMAGTYKKSDTECAACHSDCATCSSEA 203
QY	416	EPAVGFEIYKWNNLPIPTNMNTTIVLISGINFEYKGMTGWEVAGHIIYAAAGASDNDFNHLTV
Db	204	-----NNOCTSCET-----GRKL----- 216
QY	476	VPGFRPQPSVMDTEN---KEVARITVFETICLGSVNCELYFWVGNSR----- 520
Db	217	-----KSNQCVERTNCNTCNHYPDDTSMTCTSDSATAKGLCTLCNSKI 272
QY	521	-----TNTPVE-TWKGSKKGKOSTYLLENTTTSFTAORTTFAHASRKYTNDAKI 572
QY	273	PRITLDGSTCVENSYAGCQGDNEFLMKEDQSAICL-----GDTKEA SND--- 319
Db	573	YSINVNTNVWNGVASYCRPCALEASDWGSSCSPAGYYIDRSGT---CHISCPPNT- 625
QY	320	-----KGVAW-CRICKTMANDSPPTCICRLOCYFLERGSCITTCADCNTCAEATT 369
Db	626	-----ILKA-----HOPGYWQACVCPGCPETKNNKI 650
QY	370	EDKCKICKAGFFLSPGEKCC1CSQDTNNGGIDGCAECTKRPAGPLIKCTKCPNKR--- 425
Db	651	HSLCYINDCTFSRMRNTRNYYNISALATEV---TLAGSPS-----FTSKGLKFHFHET 700
QY	426	-----PAGTSD-NYIYTEKTCEDPTVCGGTGACDATVIDANGEKE---HY 467
Db	701	LSLGG---NQGRKMSVCNDN---VTDARIPGESEGF-----SKS 733
QY	468	CSTCGEINPKFPTDGLCUDKNTGAGCTDHCTSYCAAGFLYMGCYKIDTVPGSYMCMSKS 527
Db	734	ITAVCQA---VLPPEVTGKA---GVSSQPVSLADRLLGV---TTDMTL 775
QY	528	TTAGVCDTPNPNRFFVVPKAISAEOSVLACGNPLCTIAGGNAVVGVSQCTAPDARA 587
Db	776	DGITSPAELFHESLGLPDPVIFYRSNDVTOCSRSSRTIRVRCSPQTKVPGSLLPGT 835
QY	588	DGGMAYA-----TCTACEDGK-----KPGKSGTSCV---A 614
Db	836	CSDGTCDCGCHFHFLWESAACPLCSVADYHAIVSSCVAGIOTKTYVWREPKLCSQGISLP 895
Db	615	CPDANOKSCTMDVYCEE-----CADGFSLID 639
RESULT	9	9
LMLL_CAEEL	STANDARD;	PRT; 1557 AA.
QY	896	BORVTTCKTDFWLKVKGISAGTCAAILTV 925
Db	640	NGKCVSSGTTNKSGLSLTGATINGISWAIVVV 669

RP	SEQUENCE FROM N.A.
RC	STRAIN-Bristol N2;
RA	Minx P;
RL	Submitted (JAN-1996) to the EMBL/genBank/DBJ databases.
CC	CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC	-1 SIMILARITY: CONTAINS 10.5 LAMININ EGF-LIKE DOMAINS.
CC	-1 SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC	-----
CC	-----
DR	EMBL; U46673; AAC4B152.1; -.
DR	RSP; P02468; TITLE.
DR	WormPeP; C34D1..5; CBO6981.
DR	InterPro; IPR00561; EGF-like.
DR	InterPro; IPR001886; Laminin.
DR	InterPro; IPR00034; Laminin_B.
DR	InterPro; IPR002049; Laminin_EGF.
DR	Pfam; PF00052; Laminin_B; 1.
DR	Pfam; PF00053; Laminin_EGF; 10.
DR	Pfam; PF00055; Laminin_Ntmem; 1.
DR	PRINTS; PR00011; EGFLAMININ.
DR	PRODOM; PD002082; Laminin; 1.
DR	PRODOM; PD003031; Laminin_B; 1.
DR	SMART; SMO0180; EGF_Lam; 9.
DR	SMART; SMO0081; EGF-like; 2.
DR	SMART; SMO0281; Lamb; 1.
DR	SMART; SMO0136; Lamin; 1.
DR	PROSITE; PS00022; EGF_1; 8.
DR	PROSITE; PS01248; LAMININ_TYPE_EGF; 11. Hypothetical protein; Signal; Repeat.
KW	POTENTIAL. LAMININ EGF-like domain; Signal; Repeat.
FT	SIGNAL 1
FT	CHAIN ?
FT	DOMAIN ?
FT	DOMAIN 272
FT	DOMAIN 331
FT	DOMAIN 387
FT	DOMAIN 338
FT	DOMAIN 435
FT	DOMAIN 487
FT	DOMAIN 488
FT	DOMAIN 688
FT	DOMAIN 687
FT	DOMAIN 737
FT	DOMAIN 769
FT	DOMAIN 770
FT	DOMAIN 819
FT	DOMAIN 874
FT	DOMAIN 875
FT	DOMAIN 930
FT	DOMAIN 931
FT	DOMAIN 978
FT	DOMAIN 1025
FT	DISULFID 272
FT	DISULFID 274
FT	DISULFID 297
FT	DISULFID 306
FT	DISULFID 329
FT	DISULFID 341
FT	DISULFID 332
FT	DISULFID 334
FT	DISULFID 349
FT	DISULFID 360
FT	DISULFID 385
FT	DISULFID 372
FT	DISULFID 400
FT	DISULFID 406
FT	DISULFID 417
FT	DISULFID 432
FT	DISULFID 432
FT	DISULFID 449
FT	DISULFID 456
FT	DISULFID 457
FT	DISULFID 458
FT	DISULFID 467
FT	DISULFID 485
FT	DISULFID 778
FT	DISULFID 800
FT	DISULFID 817
RN	[1]

FT	DISULFID	820	834	BY SIMILARITY.
FT	DISULFID	822	841	BY SIMILARITY.
FT	DISULFID	844	853	BY SIMILARITY.
FT	DISULFID	856	872	BY SIMILARITY.
FT	DISULFID	875	894	BY SIMILARITY.
FT	DISULFID	903	912	BY SIMILARITY.
FT	DISULFID	915	928	BY SIMILARITY.
FT	DISULFID	931	943	BY SIMILARITY.
FT	DISULFID	933	950	BY SIMILARITY.
FT	DISULFID	952	961	BY SIMILARITY.
FT	DISULFID	964	976	BY SIMILARITY.
FT	DISULFID	979	991	BY SIMILARITY.
FT	DISULFID	981	998	BY SIMILARITY.
FT	DISULFID	1000	1009	BY SIMILARITY.
FT	CARBOHYD	1012	1023	BY SIMILARITY.
FT	CARBONID	101	101	BY SIMILARITY.
FT	CARBONID	121	121	BY SIMILARITY.
FT	CARBONID	121	121	BY SIMILARITY.
FT	CARBONYD	507	507	(POTENTIAL).
FT	CARBONYD	643	643	(POTENTIAL).
FT	CARBONYD	725	725	(POTENTIAL).
FT	CARBONYD	921	921	(POTENTIAL).
FT	CARBONYD	1062	1062	(POTENTIAL).
FT	CARBONYD	1168	1168	(POTENTIAL).
FT	CARBONYD	1211	1211	(POTENTIAL).
FT	CARBONYD	1244	1244	(POTENTIAL).
FT	CARBONYD	1321	1321	(POTENTIAL).
SQ	SEQUENCE	1437	1437	(POTENTIAL).
SQ	SEQUENCE	1557	AA:	CAFOB51F8D5E8D2F CRC64;
Query Match				Best Local Similarity 3 %; Score 192; DB 1; Length 1557; Matches 211; Conservative 99; Mismatches 390; Indels 394; Gaps 54;
QY	20 RRIPLRWLRLWAGTAQVQTGIGELHACKSESYHEYTACDSGCSRWRVAVFHPTGLC	79		
Db	114 QQQPPTTNTLVLGKSFDTIYVLFKTSPPRFESTTYKKTHDSEWEPHQF-----	164		
QY	80 TSLPDPKVGTECSNCNKGTSKQKLGAFASLA	139		
Db	165 -----YGSCKA-----TYGLS-----DRAILPNEATRQC	191		
QY	140 NMELDDSAAESNGCNCSKSKWTPRGDYTAFNDE-----CTATIMAVNLQKGSTVNFE--	192		
Db	192 TKEPSDISPPTGGNIASTLEGPRSPAHAFEESEVQLQKWTASA1-R-SLNRAINTFGDEV	192		
QY	193 -----YYPDSSTIFFEV-----ON-----	250		
Db	251 KDPQVRSLRSYY---AISDEAVGRCKNGHASECCVSSESSVPGENRVRCEHNTQGADC	306		
QY	209 DOCQPWADDSEWRMKYTEKGWFMHESVELN-----RCNNVAKWRTAFSVWTKVPKVLP	260		
Db	307 NECLPFYNDPRWRSGT-----SVEANACTACNCSOLSNRICYFPQOLFEE-----	350		
QY	261 VRNLTITGVAUTSBCFPCKPGTVADKGOGSSFCKLCAANSNSNKGTSCHOCDPKYSKG	320		
Db	351 -----TG---HGGHCIDQGNT---OG-VHCEQCLANHWRPGENEYCAAG--CNEIG	394		
QY	321 SSS-----CNRVRACTDK-----DFY-----THACDANGETOLMYKWAQPKI	412		
Db	395 SLSTQCDNEKGQKQCPWTGRCDCQCLDGFYDFSTNGKNCGETSGSLN---WOPRC	449		
QY	360 CSEDELGAVKLQDASGKTHCPGPNCGF -- KTMNSTQCPYGSYSNGSDCTRCP-----	456		
Db	450 DSSSGSCSKKINVEG---ROCDKCKGYFDLSTENQFCPTC--FCGHSSICNTADGYA	505		
QY	413 -----AGTERAVGFWKWNLT---PTNMNETTIVLGINFIFYKGMTCWNEAGD	555		
Db	506 MNVSVYFQDKOKWAG-QNIGLQDWTWAELKAVASYSDTONSPYF-----VAPE	456		
QY	457 HYTAGAGASNDMFLTNVPGRRPOSV---MADTENKEVARI-----T	498		
Db	556 QFLGDQRSYNQDVLFTLKVAKHTNQDVKDILVADROELSTSITAQCNPFPI TEAQ	615		
QY	499 FVFETLCSNCLETF-----MVGUNSPNTNTVETWKGSKKOSTYILEENTTS	548		
Db	616 YRFR---WHADPLFGWYPRINBLDFIGILS--NITAIRGTSYKDIGLVSNNLGT	669		
QY	549 -----FW-----AFQRTFHEASKRTTNDAKISINV	577		
Db	670 GVAPSAANKQATWIEHCECLPGFVGQFCSCESGFRRET----KFGGPENHCKDC	723		
QY	578 TNMANGVASYCRPCALEASDVGSCTSCPAGYVYDRDGT---CHSCP-PN---TILKAH	630		
Db	724 HHHSNSCEAEQSCICEHNATGDTCERGARYGVGALOGTEEDCQKCPGCPHDPC	783		
QY	631 -----QPYGVQACVCPGPKNNNHSICLYNDCTFSRTPPTTFNNNFSALANTVT	681		
Db	784 GDVICTECPMSYTGRCDBCSGDFGNPKDGTBECVACSGNTDP-----NSIGNCDK	836		
QY	682 LAG-----GPSFTSKL-----KYFHFLSLOGN	706		
Db	837 ITGECKCFCINTHGFCNCCKNCNCNGDQCAHGFYNTSGGCQCNCPDPLGSEGNTCDVNTGOCOCK	896		
QY	707 OGKMSVCUDNVDLRPEGESFSKISITAVCQAVIIP-----EV-TQY---KAG	754		
Db	897 QDGQCD-CLPNVIGIQCQDCAGFYNTSGGCQCNCPDPLGSEGNTCDVNTGOCOCK	955		
QY	755 VSSQPVSLADRLJGVTDMTDGITSPAELFHLSLGSIGDIPDVIFYRSNDVQSC--SSGR	812		
Db	956 VIGQ--RCDRADYHFGSNG--COPCOEYIGS-----ENQCDVNSG-	996		
QY	813 STMIRVRCSPQKTVPGSLLRGTCSDGTDGC--NFHFLWESAAACPLCSVADYHAIVSS	870		
Db	997 -----QCLKENVEGR-----RCDQCAENRYGITOCGLPC-----DD	1028		
QY	871 CWAGIQTKTTWRE	884		
Db	1029 CYTLIQSRVNVFRE	1042		
RESULT 10				
LMG3_HUMAN				
ID	LMG3_HUMAN	STANDARD:	PRT:	1587 AA.
QY	OYQN6;			
Db	15-JUN-2002 (Rel. 41, Created)			
QY	15-JUN-2002 (Rel. 41, Last sequence update)			
Db	15-JUN-2002 (Rel. 41, Last annotation update)			
GN	LAMC3.			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OX	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
RN	[1]			
RC	SEQUENCE FROM N_A.			
RX	MEDLINE=99242614; PubMed=10225960;			
RA	Koch M., Olson P.F., Albus A., Jin W., Hunter D.D., Brunkun W.J.,			
RA	Burgesson R.E., Chompliaud M.F.,			
RT	"Characterization and expression of the laminin gamma3 chain: a novel,			
RT	non-basement membrane-associated laminin chain,"			
RL	J. Cell Biol. 145(5-6):1818(1999).			
CC	- FUNCTION: Binding to cells via a high affinity receptor, laminin			
CC	is thought to mediate the attachment, migration, and organization			
CC	of cells into tissues during embryonic development by interacting			
CC	- SUBUNIT: Laminin 1 is a complex glycoprotein, consisting of three			
CC	different polypeptide chains (alpha, beta, gamma), which are bound			
CC	to each other by disulfide bonds into a cross-shaped molecule			
CC	comprising one long and three short arms with globules at each			
CC	- SUBCELLULAR LOCATION: Extracellular			
CC	-1- THE GAMMA-3 CHAIN IS A SUBUNIT OF LAMININ-12.			
CC	-1- TISSUE SPECIFICITY: Broadly expressed in: skin, heart, lung, and			

Wed Mar 12 10:08:37 2003

the reproductive tracts.

CC - I- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.

CC - I- DOMAIN: DOMAIN IV IS GLOBULAR.

CC - I- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).

CC - I- SIMILARITY: CONTAINS 1 LAMININ EGF-LIKE DOMAINS.

CC OR send an email to license@lsb-sib.ch).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation in the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is not removed. Usage by and for commercial modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.sib-sib.ch/announce/>).

CC EMBL; AF041835; RAB36991.1; - .

DR HSSP; P02468; ITGB.

DR Genew; HGNC; 494; LAMC3.

DR InterPro; IPR000561; EGF-like.

MIM; 604549; - .

DR InterPro; IPR001886; LamintB.

DR InterPro; IPR000034; LamminEGF.

DR InterPro; IPR002049; LamminEGF.

DR Pfam; PF00052; Lammin_B; 1.

DR Pfam; PF00053; Lammin_EGF; 10.

DR Pfam; PF00055; Lammin_Nterm; 1.

DR PRINTS; PRO0011; EGFLAMIN.

DR PRODOM; PDO02082; Lamnt_1.

DR PRODOM; PDO03031; Lammin_B; 1.

DR SMART; SM0180; EGF_Lam; 10.

DR SMART; SM00001; EGF_like; 1.

DR SMART; SM00281; LamB; 1.

DR SMART; SM00136; LamNT_1.

DR SMART; SM0022; EGF_1; 7.

DR PROSITE; PS01186; EGF_2; 2.

DR PROSITE; PS01248; LAMININ_TYPE-EGF; 10.

KW Glycoprotein; Basement membrane; Extracellular matrix; coiled coil; Laminin EGF-like domain; Cell adhesion; Repeat; Signal.

FT SIGNAL¹ 19

FT CHAIN²⁰ 1587

FT DOMAIN²⁰ 270

FT DOMAIN²⁷¹ 326

FT DOMAIN³²⁷ 382

FT DOMAIN³⁸³ 429

FT DOMAIN⁴³⁰ 479

FT DOMAIN⁴⁸⁰ 489

FT DOMAIN⁴⁹⁰ 672

FT DOMAIN⁶⁷³ 706

FT DOMAIN⁷⁵⁴ 754

FT DOMAIN⁷⁵⁵ 809

FT DOMAIN⁸¹⁰ 865

FT DOMAIN⁸⁶⁶ 916

FT DOMAIN⁹¹⁷ 964

FT DOMAIN⁹⁶⁵ 1013

FT DOMAIN¹⁰¹⁴ 1587

FT DOMAIN¹⁰⁷¹ 1141

FT DOMAIN¹²⁰⁰ 1229

FT DOMAIN¹⁴²⁴ 1504

FT DOMAIN¹⁵⁵⁵ 1579

FT SITE¹⁰⁵⁹ 1061

FT CARBOHYD⁸⁷ 87

FT CARBOHYD¹¹⁹ 119

FT CARBOHYD²⁹⁵ 295

FT CARBOHYD³²⁸ 328

FT CARBOHYD⁶³¹ 631

FT CARBOHYD⁸³⁷ 837

FT CARBOHYD⁹⁸⁰ 980

FT CARBOHYD¹¹⁸⁵ 1185

FT CARBOHYD¹⁵¹⁸ 1518

FT CARBOHYD¹⁵⁸⁷ AA;

SO SEQUENCE; MW; 3CBE609B5F203319 ; CFC64;

Db

922 LGSQEQCHPKTGTGQTCRPGVTGACDRQCLPFGSSIKGCRCRCSPLGLGASAOC 977

RESULT 11
 LMB1_HUMAN STANDARD: PRP; 1786 AA.

ID P07342; DT 01-AUG-1988 (Rel. 08, Created)
 AC R07342; DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DE 15-JUN-2002 (Rel. 41, Last annotation update)
 GN LAMBL
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90368778; PubMed=197589;
 RA Vuolteenaho R., Chow L.T., Tryggrason K.;
 RT "Structure of the human laminin Bl chain gene.";
 RL J. Biol. Chem. 265:15611-15616(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87280097; PubMed=3611077;
 RA Pikkilainen T., Eddy R., Fukushima Y., Byers M., Shows T.,
 RT "Human laminin Bl chain. A multidomain protein with gene (LAMBL)
 locus in the q22 region of chromosome 7.";
 RL J. Biol. Chem. 262:10454-10462(1987).
 RN [3]
 RP SEQUENCE OF 1276-1909 FROM N.A.
 RX MEDLINE=8803102; PubMed=361559;
 RA Jaye M., Modis W.S., Ricca G.A., Mudd R., Chiu T.M., O'Brien S.J.,
 RT Drohan W.N.;
 RT "Isolation of a cDNA clone for the human laminin-Bl chain and its
 gene localization.";
 RL Am. J. Hum. Genet. 41:605-615(1987).
 CC -!- FUNCTION: Binding to cells via a high affinity receptor, laminin
 CC is thought to mediate the attachment, migration, and organization
 CC of cells into tissues during embryonic development by interacting
 CC with other extracellular matrix components.
 CC -!- SUBUNIT: laminin is a complex glycoprotein, consisting of three
 CC different polypeptide chains (alpha, beta, gamma), which are bound
 CC to each other by disulfide bonds into a cross-shaped molecule
 CC comprising one long and three short arms with globules at each
 CC end.
 CC THE BETA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ), LAMININ-
 CC 2 (HEROSIN), AND LAMININ-5 (K-LAMININ).
 CC -!- SUBCELLULAR LOCATION: extracellular.
 CC -!- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
 CC -!- COMPONENT).
 CC -!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
 CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
 CC -!- DOMAIN: DOMAINS VI AND IV ARE GLLOBULAR.
 CC -!- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
 CC -!- SIMILARITY: CONTAINS 12.5 LAMININ EGF-LIKE DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@ib-sib.ch).

DR EMBL; M61923; AAA59486.1; JOINED.
 DR EMBL; M61924; AAA59486.1; JOINED.
 DR EMBL; M61925; AAA59486.1; JOINED.
 DR EMBL; M61926; AAA59486.1; JOINED.
 DR EMBL; M61927; AAA59486.1; JOINED.
 DR EMBL; M61928; AAA59486.1; JOINED.
 DR EMBL; M61929; AAA59486.1; JOINED.
 DR EMBL; M61930; AAA59486.1; JOINED.
 DR EMBL; M61931; AAA59486.1; JOINED.
 DR EMBL; M61932; AAA59486.1; JOINED.
 DR EMBL; M61933; AAA59486.1; JOINED.
 DR EMBL; M61934; AAA59486.1; JOINED.
 DR EMBL; M61935; AAA59486.1; JOINED.
 DR EMBL; M61936; AAA59486.1; JOINED.
 DR EMBL; M61937; AAA59486.1; JOINED.
 DR EMBL; M61938; AAA59486.1; JOINED.
 DR EMBL; M61939; AAA59486.1; JOINED.
 DR EMBL; M61940; AAA59486.1; JOINED.
 DR EMBL; M61941; AAA59486.1; JOINED.
 DR EMBL; M61942; AAA59486.1; JOINED.
 DR EMBL; M61943; AAA59486.1; JOINED.
 DR EMBL; M61944; AAA59486.1; JOINED.
 DR EMBL; M61945; AAA59486.1; JOINED.
 DR EMBL; M61946; AAA59486.1; JOINED.
 DR EMBL; M61947; AAA59486.1; JOINED.
 DR EMBL; M61948; AAA59486.1; JOINED.
 DR EMBL; M61949; AAA59486.1; JOINED.
 DR EMBL; M61950; AAA59486.1; JOINED.
 DR EMBL; M61944; AAA59486.1; JOINED.
 DR EMBL; M61945; AAA59486.1; JOINED.
 DR EMBL; M61946; AAA59486.1; JOINED.
 DR EMBL; M61947; AAA59486.1; JOINED.
 DR EMBL; M61948; AAA59486.1; JOINED.
 DR EMBL; M61949; AAA59486.1; JOINED.
 DR EMBL; M53365; AAA59485.1; JOINED.
 DR EMBL; M53371; AAA59485.1; JOINED.
 DR EMBL; M53372; AAA59485.1; JOINED.
 DR EMBL; M53373; AAA59485.1; JOINED.
 DR EMBL; M53374; AAA59485.1; JOINED.
 DR EMBL; M53375; AAA59485.1; JOINED.
 DR EMBL; M53376; AAA59485.1; JOINED.
 DR EMBL; M53377; AAA59485.1; JOINED.
 DR EMBL; M53378; AAA59485.1; JOINED.
 DR EMBL; M53379; AAA59485.1; JOINED.
 DR EMBL; M53380; AAA59485.1; JOINED.
 DR EMBL; M53381; AAA59485.1; JOINED.
 DR EMBL; M53382; AAA59485.1; JOINED.
 DR EMBL; M53383; AAA59485.1; JOINED.
 DR EMBL; M53384; AAA59485.1; JOINED.
 DR EMBL; M53385; AAA59485.1; JOINED.
 DR EMBL; M53386; AAA59485.1; JOINED.
 DR EMBL; M53387; AAA59485.1; JOINED.
 DR EMBL; M53388; AAA59485.1; JOINED.
 DR EMBL; M53389; AAA59485.1; JOINED.
 DR EMBL; M53390; AAA59485.1; JOINED.
 DR EMBL; M53391; AAA59485.1; JOINED.
 DR EMBL; M53392; AAA59485.1; JOINED.
 DR EMBL; M53393; AAA59485.1; JOINED.
 DR EMBL; M53394; AAA59485.1; JOINED.
 DR EMBL; M53395; AAA59485.1; JOINED.
 DR EMBL; M53396; AAA59485.1; JOINED.
 DR EMBL; M53397; AAA59485.1; JOINED.
 DR EMBL; M53398; AAA59485.1; JOINED.
 DR EMBL; M53399; AAA59485.1; JOINED.
 DR EMBL; M53400; AAA59485.1; JOINED.
 DR EMBL; M53361; AAA59485.1; JOINED.
 DR EMBL; M53362; AAA59485.1; JOINED.
 DR EMBL; M53363; AAA59485.1; JOINED.
 DR EMBL; M53364; AAA59485.1; JOINED.
 DR EMBL; M53365; AAA59485.1; JOINED.
 DR EMBL; M53366; AAA59485.1; JOINED.
 DR EMBL; M53367; AAA59485.1; JOINED.
 DR EMBL; M53368; AAA59485.1; JOINED.
 DR EMBL; M53369; AAA59485.1; JOINED.
 DR EMBL; M61916; AAA59482.1; JOINED.
 DR EMBL; M20206; AAA59487.1; JOINED.
 DR EMBL; M53347; MMHUOL.
 DR HSSP; P03468; IKIQL.
 DR Gene; HGNC:6486; LAMBL.
 DR MIM: 150240; JOINED.
 DR InterPro; IPR00056; EGF-like.
 DR InterPro; IPR001886; Lamnt.
 DR Pfam; PF00053; Laminin_EGF; 13.
 DR Pfam; PF00055; Laminin_NFerm; 1.
 DR PRINTS; PRO0011; EGFLAMININ.
 DR ProDom; PD002082; Lamnt; 1.

Wed Mar 12 10:08:37 2003

RA STRAINC-57BL/6; TISSUE-Thymus;
RX MEDLINE-93346725; PubMed=8345183;
RA Chang A.C.; Wadsworth S.; Colligan J.E.;
RT "Expression of merosin in the thymus and its interaction with
trophocytes";
RT J. Immunol. 151:1789-1801(1993).
RN [3]
RN SEQUENCE OF 64-281 FROM N.A.
RP MEDLINE-95179178; PubMed=787173;
RA Xu H.; Wu X.R.; Wever U.M.; Engvall E.;
RT "Murine muscular dystrophy caused by a mutation in the laminin alpha
2 (Lama2) gene"; Nat. Genet. 8:297-302(1994).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.35 ANGSTROMS) OF 2932-3106.
RA Hohenester E.; Tisi D.; Taits J.F.; Timpl R.;
RT "The crystal structure of a laminin G-like module reveals the
molecular basis of alpha-dystroglycan binding to laminins, perlecan,
and agrin"; Mol. Cell 4:783-792(1999).
RL -!- FUNCTION: Binding to cells via a high affinity receptor, laminin
is thought to mediate the attachment, migration, and organization
of cells into tissues during embryonic development by interacting
with other extracellular matrix components.
CC -!- SUSUNIT: Laminin is a complex glycoprotein, consisting of three
different polypeptide chains (alpha, beta, gamma), which are bounded
to each other by disulfide bonds into a cross-shaped molecule
COMPONENT).
CC Comprising one long and three short arms with globules at each
end.
CC THE ALPHA-2 CHAIN IS A SUBUNIT OF LAMININ-2 (MEROSEN) AND LAMININ-1.
CC 4 (S-MEROSIN).
CC -!- DOMAIN: Laminin is a complex glycoprotein, consisting of three
different polypeptide chains (alpha, beta, gamma), which are bounded
to each other by disulfide bonds into a cross-shaped molecule
COMPONENT).
CC -!- TISSUE_SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
COMPONENT).
CC -!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
CC -!- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.
CC -!- DISEASE: DEFECTS IN LAMAZ ARE A CAUSE OF MURINE MUSCULAR DYSTROPHY
(DMD).
CC -!- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC -!- SIMILARITY: CONTAINS 17 LAMININ EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.
CC -!- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce.html>).
CC
EMBL: U12147; AAC2165_1; -;
EMBL: S75315; AAB3573_1; -;
PDB: 1Q00; 03-DEC-99;
MGD: MGI:99912; Lam2.
InterPro: IPR000561; EGF-like.
InterPro: IPR001886; LamNT.
InterPro: IPR000034; Laminin_B.
InterPro: IPR002049; Laminin_BQF.
InterPro: IPR001791; Laminin_G.
PFam: PF00052; Laminin_B.
PFam: PF00053; Laminin_EGF; 15.
PFam: PF00054; Laminin_G; 5.
PFam: PF00055; Laminin_Nterm; 1.
PRINTS: PRO0011; EGFLAMININ.
PRODom: PD002082; LamMT; 1.
SMART: SM00180; EGF_Lam; 15.
SMART: SM00001; EGF_like; 1.
SMART: SM00281; LamB; 2.

FT DISULFID 965 981 BY SIMILARITY.
 FT DISULFID 983 992 BY SIMILARITY.
 FT DISULFID 995 1007 BY SIMILARITY.
 FT DISULFID 1010 1019 BY SIMILARITY.
 FT DISULFID 1012 1026 BY SIMILARITY.
 FT DISULFID 1028 1037 BY SIMILARITY.
 FT DISULFID 1040 1053 BY SIMILARITY.
 FT DISULFID 1056 1068 BY SIMILARITY.
 FT DISULFID 1058 1075 BY SIMILARITY.
 FT DISULFID 1077 1086 BY SIMILARITY.
 FT DISULFID 1089 1099 BY SIMILARITY.
 FT DISULFID 1416 1425 BY SIMILARITY.
 FT DISULFID 1418 1432 BY SIMILARITY.
 FT DISULFID 1435 1444 BY SIMILARITY.
 FT DISULFID 1447 1462 BY SIMILARITY.
 FT DISULFID 1465 1480 BY SIMILARITY.
 FT DISULFID 1467 1490 BY SIMILARITY.
 FT DISULFID 1493 1502 BY SIMILARITY.
 FT DISULFID 1505 1520 BY SIMILARITY.
 FT DISULFID 1523 1535 BY SIMILARITY.
 FT DISULFID 1525 1542 BY SIMILARITY.
 FT DISULFID 1544 1553 BY SIMILARITY.
 FT DISULFID 1556 1567 BY SIMILARITY.
 FT DISULFID 1570 1570 INTERCHAIN (PROBABLE).
 FT DISULFID 1574 1574 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 51 51 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 85 85 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 359 359 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 376 376 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 466 466 N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Matchn 3.1%; Score 172.5; DB 1; Length 3106;

Best Local Similarity 19.5%; Pred. No. 0.00047; Indels 403; Gaps 57; Matches 191; Conservative 80; Mismatches 304;

QY 42 TGBELHACKESEY------HWEYT---ACDSTGSRMWRVAVPHPTG---- 77
 836 TGRERCERAGEGYFGQSPVPGGSCQPCQNDLNDIYLPGSCDSLGSCLICKPTGTYRGE 895

QY 78 LCMS--LPPVKGTEC-----SFS--CM--AGEFLDKMDKQSKPKAEGRRY----- 116
 Db 896 LCADGYFGDAVNWKNCQPCRCLINGSEDCRTTGTGC-----ECRPNVGRHDECKP 949

QY 117 -----SLGTGIRDEWDELPLP--HGFAISLAMELDDSAESTNGNTSSKAWPRGDYIAF 168

Db 950 ETPGLQLGRGC-----LPCNNSFGSKSF-----CEASQC----WCQPG-VAGK 990

Qy 169 NTPECTATLMAYVNLKGSGTVFEEVYYPDSSIFEFFVQNDQCPNADDSRWMKITEKGW 228

Db 991 KCBRCA-----HGYNHQ----EGGCIACDCSHLGNND-----KT----- 1023

Qy 229 EPHISVELNRGNNNLYIWR-TAFSWTKVPKPVLRVMIAITGVAYTSECPCKPGTVADKQG 288

Db 1024 -----GQCCPPNITG-----EKSECPLNTW---G 1046

Qy 289 SSF---CKLCPANSYS-----NKGETSCHOCDPDKYSEKGSSSN---VRPACTDK 333

Db 1047 HSTVTGCKVCNCSTVGSLASQCNWTGQCSCH---PKFSGMKCSECSRGRHWNPICTLC 1102

Qy 334 DYFY---IHTACDANGEROLMMFKWAKPKICSDELEG--AVKUPLASGWYKTHCOPPCNQGFF- 387

Db 1103 DCFLPGLGTDATTCTCL---ETR-----KCSCSDQTCGQCSCKVNVNEGV--HCDRCRGKFG 1150

Qy 388 -KTNINNTOQPCYGYSNS--SDCTRCPAGEIPEAVGSEYKWWNLTPTNEMTTVLGINF 443

Db 1151 LDAKNPLGCCSC---YFGVHSQSCSE-----AKGL-IRWVTL--SDPQTILPV-- 1194

Qy 444 EYKGMTGWVAGDIHYTAGASINDEMILTLVVPGFRPQSMADTENKEYARI----- 497

Db 1195 -----DEALQHTTKGIA-----FQKEPEVAKMDEVROELHLEPFWK 1232

Qy 498 --TFVFETLCVNCELYFMVGVNSRTNPVETWKGS---RKQSYTYIIEENTITSFT 550

RESULT 13
 TS1L_GITALA ID TS1L_GITALA STANDARD; PRT; 667 AA.
 AC 003185;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Major surface trophozoite antigen 11 precursor.
 GN TSPL.
 OS Giardia lamblia (Giardia intestinalis).
 OC Eukaryota: Diplomonadida; Hexamitidae; Giardinae; Giardia.
 OX NCBI_Taxid=5741;
 RN
 RP SEQUENCE FROM N A.
 RC STRAIN=Isolate AD-1;
 RX MEDLINE=9341215; PubMed=8479449;
 RA EV P. L., Khanna K., Manning P.A., Mayrhofer G.;
 RT "A gene encoding a 69 kilodalton major surface protein of Giardia
 intestinalis trophozoites.";
 RL Mol. Biochem. Parasitol. 58:47-258(1993).
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ANCHORED TO THE
 CC PLASMA MEMBRANE.
 CC -!- DOMAIN: CONTAINS 27 REPEATS OF A CXIC MOTIF.
 CC -!- SIMILARITY: BELONGS TO THE GIARDIA VARIANT SURFACE PROTEIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to licensee@lsb-sib.ch).
 CC
 DR EMBL; M95814; AAA02687.1; -.
 DR PIR; A48579; A8579.
 DR HSSP; P02468; IKLO.
 DR InterPro; IPR00561; EGF-like.
 DR InterPro; IPR00174; Furan-like.
 DR InterPro; IPR05127; Giardia_vsp.
 DR Pfam; PF0302; VSP; 1.
 DR SMART; SM00181; EGF; 3.

DR	KW SMART; S400261; FU: 5.	RN [1]
FT	Antigen; Repeat; Transmembrane; Signal.	SEQUENCE FROM N.A.
FT	SIGNAL 1 17	RX MEDLINE=8714722; PubMed=3493487;
FT	CHAIN 18 667	RX Sasaki M., Kato S., Kohno K., Martin G.-R., Yamada Y.;
FT	DOMAIN 18 633	"Sequence of the cDNA encoding the laminin BI chain reveals a
FT	TRANSMEM 634 662	POTENTIAL; multidomain protein containing cysteine-rich repeats";
FT	DOMAIN 663 667	CYTOSLASMIC (POTENTIAL).
FT	CARBHYD 591 591	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBHYD 630 630	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE 667 AA; 68475 MW; 1DD95727@3232B8D CRC64;	RX Barlow D.P., Green N.M., Kurkinen M., Hogan B.L.M.;
		"Sequencing of laminin B chain cDNAs reveals C-terminal regions of
		coiled-coil alpha-helix";
		EMBO J. 3:2355-2362(1984).
QY	91 CSFSCNAGAEFLDMMDOSCKPCABG-RYSIGTGFIDELPHFASLUSANMELDDSAE 149	RN [3]
Db	141 CSARKAPKGEDNTPKPAATKCAAGFLHPTSEGSISSCE ETCPEGYFGHA-----"NAE 192	SEQUENCE OF 165-172: 539-547 AND 712-719.
Qy	150 STGNCCTSSKKWVPRDYIAFTDECTATIMAVNIKQSGTVNFEEYYPDSIIEFFVOND 209	RN STRAIN=BALB/C; TISSUE=Endothelial cells;
Db	193 SKKICKS-----CTGGSEAPWNKGIGDCLKNNYNEASGNTLCEKSA 236	RX MEDLINE=97363207; PUBMED=9219532;
Qy	210 QCOPNADDSRWKMKTEKSGMEFHSLVNLRNRRNVLYWRTAFSWMVTKVPKVLVRNIAITGV 269	RX Frieser M., Noeckel H., Pausch F., Roeder C., Hahn A., Deutzmann R., Sorokin I.M.;
Db	237 QKPKSLDKTSCNQCT-----GONCAFESSG----- 262	RX "Cloning of the mouse laminin alpha 4 cDNA. Expression in a subset of endothelium."
Qy	270 AYTSECFCRKPGTYADKOG-----SSFKLICPANSYSNKGEPS--HQCDDPKYSEK 319	RX Eur. J. Biochem. 246:727-735(1997).
Db	263 ---GDCEGEGDSGTILDGQNCVKSDCKTENKACINPKAANEVCIECLSTHHLTP----- 313	RX -1- FUNCTION: Binding to cells via a high affinity receptor. Laminin is thought to mediate the attachment, migration, and organization of cells into tissues during embryonic development by interacting with other extracellular matrix components.
Qy	320 GSSSCNVRPACTDKDVYTHTACDANGETQLMYKKWAKPKICSDELEGAVKLPGVKI----- 377	CC -1- SUBUNIT: Laminin is a complex glycoprotein, consisting of three different polypeptide chains (alpha, beta, gamma), which are bound to each other by disulfide bonds into a cross-shaped molecule.
Db	314 -TSQC---VOYCQALGNVYAGTNAD-----"NPKACE-----CTVANCKTCN 351	CC Comprising one long and three short arms with globules at each end.
Qy	378 --HCPPCNPNGFFKTTNNSTCOPCGY-----SYNSNGSDCTRCPPGTEPAGVFEKWWNTLP 430	CC THE BETA-1 CHAIN IS A SUBUNIT OF LAMINTIN-1 (EHS LAMININ), LAMININ-2 (MEROSIN), AND LAMININ-6 (K-LAMININ).
Db	352 DQGCQQTNDGFYK-NGDACSPCHESCKTCAGTASDCTECPIG----- 394	CC -1- SUBCELLULAR LOCATION: Extracellular.
Qy	431 TNMFTVLSGINFRNEYKGMGWEAVGHDHYTAAGASDNDMFITIVRGRF-----RP 481.	CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
Db	395 -----KALKYGDNGTKG-TCGEGCTTGQGSACKTCGLIDGASYCSEDTQNEYP 444	CC COMPONENT: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
Qy	482 PQSMADPENKEVA-RITFVEETLCSVNCLYFM-----VGVNSRT 521	CC -1- SIMILARITY: CONTAINS 12.5 LAMININ EGF-LIKE DOMAINS.
Db	445 QNGCTSTTARTVATCKNSNVANGICS-SCTNGFLRMNGCYETKFPGSKVCEGANADA 503	CC -1- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.
Qy	522 NP---PVETWKGSKKGQSYTYIILENITTSFTWAQRTFHEASRKITNDVAKIYISINT 578	CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
Db	504 DTCKAPVPGKVYEAKG--LVMCSKGCDT---CSDATCTKGGGY-----KLENS 549	CC -----
Db	579 NVWNGVASYCRPGALEASDVGSSCTSCDAGY-IDRDSGTCHS CPPNLTIKAHQPYQVA 637	DR EMBL; X05212; CAA28839.1;
Db	550 QTCCKCDSCTC---TGAATICKVCGTGYKTALEGESTCTSCEND---SNGVIGVK 601	DR PIR; P26413; MMWSBL.
Qy	638 CVPCCPGPKNNKIHSLCY 655	DR HSSP; P02468; KLO; MGD; MGI: 96743; LAMB-1.
Db	602 CLNCAPSSSTG-SVLCY 618	DR InterPro; IPR00186; LamNt.
	RESULT 14	DR InterPro; IPR00186; LamNt.
LAMB1_MOUSE	STANDARD; PRT; 1786 AA.	DR InterPro; IPR002049; Laminin_EGF.
ID	LAMB1_MOUSE STANDARD; PRT; 1786 AA.	DR Pfam; PF00053; Laminin_EGF; 13.
AC	P02469; 21-JUL-1986 (Rel. 01, Created)	DR Pfam; PF00055; Laminin_Ntter; 1.
DT	01-JUL-1989 (Rel. 11, Last sequence update)	DR PRINTS; PRO001; EGF_LAMININ.
DT	15-JUN-2002 (Rel. 41, last annotation update)	DR ProDom; PD02082; LamNt; 1.
DT	Laminin beta-1 chain precursor (Laminin B1 chain).	DR SMART; SM00180; EGF_Lam; 11.
DE	LAMB1-1 OR LAMB-1.	DR SMART; SM00136; LamNt; 1.
GN	MUS musculus (Mouse).	DR PROSITE; PS00022; EGF_1..9.
OS	Chordata; Craniata; Vertebrata; Euteleostomi;	DR PROSITE; PS01186; EGF_2..2.
OC	Eukaryota; Metazoa; Mammalia; Eutheria; Rodentia; Säugetiere; Muridae; Murinae; Mus.	DR PROSITE; PS01248; LAMININ_IYPE_ESF; 11.
OX	NCBI_TAXID=10090;	FT Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil; SIGNAL 1 21 LAMININ BETA-1 CHAIN.
		FT CHAIN 22 1786 LAMININ N-TERMINAL (DOMAIN VI).
		FT DOMAIN 22 270 LAMININ EGF-LIKE 1.
		FT DOMAIN 271 334 LAMININ EGF-LIKE 2.

FT	DOMAIN	398	457	LAMININ EGFP-LIKE 3.	FT	CARBOHYD	1343	1343	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	DOMAIN	458	509	LAMININ EGFP-LIKE 4.	FT	CARBOHYD	1487	1487	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	DOMAIN	510	540	LAMININ EGFP-LIKE 5 (INCOMPLETE).	FT	CARBOHYD	1533	1533	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	DOMAIN	541	772	LAMININ DOMAIN IV.	FT	CARBOHYD	1542	1542	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	DOMAIN	773	820	LAMININ EGFP-LIKE 6.	FT	CARBOHYD	1643	1643	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	DOMAIN	821	866	LAMININ EGFP-LIKE 7.	FT	CONFFLICT	1531	1534	SIGNA -> MEMP (IN REF. 2).
FT	DOMAIN	867	916	LAMININ EGFP-LIKE 8.	FT	CONFFLICT	1749	1749	D -> N (IN REF. 2).
FT	DOMAIN	917	975	LAMININ EGFP-LIKE 9.	FT	SEQUENCE	1786	AA;	196904 MW;
FT	DOMAIN	976	1027	LAMININ EGFP-LIKE 10.	FT	SEQUENCE	1786	AA;	846671B7BF41A474 CRC64;
FT	DOMAIN	1028	1083	LAMININ EGFP-LIKE 11.					
FT	DOMAIN	1084	1131	LAMININ EGFP-LIKE 12.					
FT	DOMAIN	1132	1178	LAMININ EGFP-LIKE 13.					
FT	DOMAIN	1179	1397	DOMAIN II.					
FT	DOMAIN	1398	1430	DOMAIN ALPHA.					
FT	DOMAIN	1431	1471	DOMAIN I.					
FT	DOMAIN	1216	1315	COILED COIL (POTENTIAL).	QY	-HOCD-----	275	CFCPKPGTY---ADKQGSFCKLC---PANS-----YSN-----KGTS---	308
FT	DOMAIN	1368	1388	COILED COIL (POTENTIAL).	Db	374	CEOCKPFYVQHPERDIRDPLNCEPCTCDPAGSENGICGICGTDPSVGLIAGQCRLKHE	433	
FT	DOMAIN	1448	1778	COILED COIL (POTENTIAL).	QY	309 -HOCD-----	275	CFCPKPGTY---ADKQGSFCKLC---PANS-----YSN-----KGTS---	308
FT	DISULFID	271	280	BY SIMILARITY.	Db	434	GERCIVDCKEGFYDLISAEDPYGCK--SCACHPLGTIPGGNPQCDSETGTCYCYKRLVTGQRCD-	491	
FT	DISULFID	273	298	BY SIMILARITY.	QY	345 NGETQLMYKWAKPKICSEDELAGVKLPSAVKTHCPPCNCNGFFKTNNSCQ---PCPKG	400		
FT	DISULFID	300	309	BY SIMILARITY.	Db	492 -->CLPQWG---LSNLQDG-----CRPCDCDGLGALNNCSEDSGQSCS	532		
FT	DISULFID	312	332	BY SIMILARITY.	QY	401 SYNSGSDCT-----	275	CFCPKPGTY---ADKQGSFCKLC---PANS-----YSN-----KGTS---	308
FT	DISULFID	335	344	BY SIMILARITY.	Db	533 PHM1GRQCNEVESGYFTTLDHYTAEANLGLFGVVVERQYQDRIFSWTG--GFWR	590		
FT	DISULFID	337	362	BY SIMILARITY.	QY	422 -----EKWVNNTLPTNMPETVSGINFYKGMTGWEYAGDHLYTAACASDNDFMLLT	474		
FT	DISULFID	365	374	BY SIMILARITY.	Db	591 VPEGAYLEF-FIDNIPYSMEYEL--IRYEPOLPDHWEKA-----	631		
FT	DISULFID	377	395	BY SIMILARITY.	QY	475 WPEFRPPQS---VMADENKEV-----ARITVFETCSVNCLEYFVQVGUN--SRNT	523		
FT	DISULFID	398	411	BY SIMILARITY.	Db	632 QRPQKTPASSRCGNVPPDDQNVVSLSPSSRYVLPVPYC-----FBRGMNTVRL	684		
FT	DISULFID	400	426	BY SIMILARITY.	QY	524 PVETWKGSKQKOSVYIE-----	275	CFCPKPGTY---ADKQGSFCKLC---PANS-----YSN-----KGTS---	308
FT	DISULFID	428	437	BY SIMILARITY.	Db	685 PQYFASGSPVESPVYFIDSJLVLMPYCKSLIDIFTVGSGGEVNTSAWFQYRCLENS	744		
FT	DISULFID	440	455	BY SIMILARITY.	QY	564 KYT---NDOKA--IYSINTNTNVNGVASYCRP-----	275	CFCPKPGTY---ADKQGSFCKLC---PANS-----YSN-----KGTS---	308
FT	DISULFID	458	472	BY SIMILARITY.	Db	692 GLKTFHHFTSLCGNQGRKMSVCTDNTVTLRIPBG-ESG--FSKSI-----	603		
FT	DISULFID	460	479	BY SIMILARITY.	QY	604 SCPAG-----	275	CFCPKPGTY---ADKQGSFCKLC---PANS-----YSN-----KGTS---	308
FT	DISULFID	481	490	BY SIMILARITY.	Db	745 SVVKTPMTDVCRNLLFISALIHOTGLACBCDPQGLSSVCDPNGQOCOCRPNVGRIC	631		
FT	DISULFID	493	507	BY SIMILARITY.	QY	805 RCARGTFFGPGNGKPCDCDHLQGSASAFAQDAITQCHCQIYARQCDKLPGY-----	858		
FT	DISULFID	517	785	BY SIMILARITY.	QY	632 PYGVQACVPOGPQTKNNKIHSLCYNDCTSRSNTPTRTFNINFSALANTVLAGGSPFTSK	691		
FT	DISULFID	773	803	BY SIMILARITY.	Db	859 -WGPSPSCQC-----QCNASHAL---DC-----DTW-----GECLSCQ	887		
FT	DISULFID	844	851	BY SIMILARITY.	QY	692 GLKTFHHFTSLCGNQGRKMSVCTDNTVTLRIPBG-ESG--FSKSI-----	740		
FT	DISULFID	864	876	BY SIMILARITY.	Db	888 DYTGFHNCERCLAGYGDPIPGSDGHCRCPCPGDPDSQRFAKSCYQDPVQLACVD	947		
FT	DISULFID	866	883	BY SIMILARITY.	QY	741 AVLIPPEVYKAGVSSQPVSLADRUGHTMDLGITSPAELFLESLIGIPDV-----	795		
FT	DISULFID	886	895	BY SIMILARITY.	Db	943 -----GQY--IGSRDDCAGGGFQPSD--FGGSCQHCNNIDTPDACKDT	994		
FT	DISULFID	898	914	BY SIMILARITY.	QY	796 -----IFFYRSNDVYQSCSG-----RESTTRVRCSPQKTVP---GSIL---LPGTC	836		
FT	DISULFID	917	933	BY SIMILARITY.	Db	995 GRCLUKLYTEGDHQLCQCGYQYGDALRDRDKCVCNYLGTVKHNGSDCHCDKATGQ	1054		
FT	DISULFID	944	955	BY SIMILARITY.	QY	837 S-----D3TCDGCFNFHFLWE--SAAACLCSVADYHAIVSSC	871		
FT	DISULFID	958	973	BY SIMILARITY.	Db	1055 SCLPNWIGONCDRCAPN-TWOLASGTCGCPNCNAHHSFGPSC	1096		
FT	DISULFID	976	990	BY SIMILARITY.					
FT	DISULFID	978	997	BY SIMILARITY.					
FT	DISULFID	1000	1009	BY SIMILARITY.					
FT	DISULFID	1012	1025	BY SIMILARITY.					
FT	DISULFID	1084	1096	BY SIMILARITY.					
FT	DISULFID	1086	1103	BY SIMILARITY.					
FT	DISULFID	1105	1114	BY SIMILARITY.					
FT	DISULFID	1117	1129	BY SIMILARITY.					
FT	DISULFID	1132	1144	BY SIMILARITY.					
FT	DISULFID	1153	1162	BY SIMILARITY.					
FT	DISULFID	1165	1176	BY SIMILARITY.					
FT	INTERCHAIN (PROBABLE).	1179	1179	INTERCHAIN (PROBABLE).					
FT	INTERCHAIN (PROBABLE).	1182	1182	INTERCHAIN (PROBABLE).					
FT	INTERCHAIN (PROBABLE).	1785	1785	INTERCHAIN (PROBABLE).					
FT	N-LINKED (GLCNAC. . .) (POTENTIAL).	120	120	N-LINKED (GLCNAC. . .) (POTENTIAL).					
FT	CARBOHYD	356	519	N-LINKED (GLCNAC. . .) (POTENTIAL).					
FT	CARBOHYD	519	677	N-LINKED (GLCNAC. . .) (POTENTIAL).					
FT	CARBOHYD	1041	1195	N-LINKED (GLCNAC. . .) (POTENTIAL).					
FT	CARBOHYD	1279	1279	N-LINKED (GLCNAC. . .) (POTENTIAL).					
FT	CARBOHYD	1336	1336	N-LINKED (GLCNAC. . .) (POTENTIAL).					
				RESULT 15					
				FUR2_DROME					
				ID	STANDARD;	PRT:	1680 AA.		

AC P30432;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Furin-like protease 2 precursor (EC 3.4.21.75) (Furin 2).
 GN OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephidoidea; Drosophilidae; Drosophila.
 NCBI-TaxID:7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=2381036; PubMed=1512259;
 RA Roebroek A.J.M., Creemers J.W.M., Pauli T.G.L., Kurzik-Dumke U.,
 Rentrop M., Gatteff E.A.F., Leunissen J.A.M., van de Ven W.J.,
 Cloning and functional expression of Dfurin2, a subtilisin-like
 RT proprotein processing enzyme of Drosophila melanogaster with multiple
 RT repeats of a cysteine motif.;
 RL J. Biol. Chem. 267:17208-17215 (1992).
 --! FUNCTION: FURIN IS LIKELY TO REPRESENT THE UNIQUitous ENDOPROTEASE
 CC ACTIVITY WITHIN CONSTITUTIVE SECRETORY PATHWAYS AND CAPABLE OF
 CC CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF (BY SIMILARITY).
 CC -! CATALYTIC ACTIVITY: Release of mature proteins from their
 CC proproteins by cleavage of Arg-xaa-Yaa-Arg-(zaa bonds, where xaa
 CC can be any amino acid and Yaa is Arg or Lys. Releases albumin,
 CC complement component C3 and von Willebrand factor from their
 CC respective precursors.
 CC -! SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8. FURIN SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 EMBL: M94375; AAA28551.1; --.
 DR PIR: A43434; A43434.
 DR HSSP: Q99405; IMPT.
 DR MEROPS: 508.049; --.
 DR Flybase: FBgn0004598; Fur2.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR002884; P domain.
 DR InterPro: IPR000209; Peptidase_S8.
 DR Pfam: PF00082; Peptidase_S8; 1.
 DR Pfam: PF01483; P; PARTIAL.
 DR PRINTS: PRO0723; SUBTILISIN.
 DR ProDom: PDO000717; P-domain; 1.
 DR SMART: SM00261; EGF; 1.
 DR PROSITE: PS00136; SUBTILASE_HIS; 1.
 DR PROSITE: PS00138; SUBTILASE_SER; 1.
 KW Hydrolase; Serine protease; Glycoprotein; Signal; Transmembrane;
 KW Multigene family; Zymogen; Repeat.
 FT SIGNAL 1 POTENTIAL.
 FT PROPEP ? 319 POTENTIAL.
 FT CHAIN 320 1680 FURIN-LIKE PROTEASE 2.
 FT ACT_SITE 418 418 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 457 457 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 638 638 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DOMAIN 962 1447 10 X TANDEM REPEATS, CYS-RICH.
 FT REPEAT 962 1007 1.
 FT REPEAT 1008 1057 2.
 FT REPEAT 1058 1104 3.
 FT REPEAT 1105 1153 4.
 FT REPEAT 1154 1205 5.
 FT REPEAT 1205 1254 6.
 FT REPEAT 1255 1299 7.
 FT REPEAT 1300 1346 8.

FT REPEAT 1347 1393 9.
 FT REPEAT 1394 1444 10.
 FT TRANSMEM POTENTIAL.
 FT DOMATN CYTOPASMIC (POTENTIAL).
 FT CARBOHYD N-LINED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 109 109 N-LINED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 130 130 N-LINED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 203 203 N-LINED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 443 443 N-LINED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 481 481 N-LINED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 928 928 N-LINED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1061 1061 N-LINED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1182 1182 N-LINED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1275 1275 N-LINED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1278 1278 N-LINED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1440 1440 N-LINED (GLCNAC. . .) (POTENTIAL).
 SQ 1680 AA; 183599 MW; 0A99CE8770A8E293 CRC64;

Query Match 3.1%; Score 169; DB 1; Length 1680;
 Best Local Similarity 17.2%; Pred. No. 0.0038; Gaps 41;
 Matches 159; Conservative 85; Mismatches 291; Indels 388; Gaps 41;

QY 46 LHACKESEYHVEYTCADSTG- SRRWV-----APVPHPELCTS 82
 Db 965 LHSCD-----ABCDSCSGCYGRGPTQCVACSHYRLNDTCVSRCRPPSPFPNQVGICWC 1016
 QY 83 PDPVKGTCCFSCNACEFLMDKQDOSKPCBEGR--YSLGIGIRDEWDDELPHGAFSLA 139
 Db 1017 HD----TCEFCAGAG----PNADDSSRMKTEKGHEFHVSVELNRNNVLWRRTA 248
 QY 140 NMELDDSAAESSTGNNTSSKPKVPRGYDIAFNNDCTATMLYAVNLQSGTINFEVYYPDS 199
 Db 1061 NR---TCVPCPNCASCODHP-----EYCTSC-----DHHLVMEHKYC 1099
 QY 200 LIFEFFVQNDQCQ-----PNADDSSRMKTEKGHEFHVSVELNRNNVLWRRTA 248
 Db 1100 PLDTYETEDNIKCAFCHSTCATCNGPDDCCTCRSSRYAQ-----NKCLSCPDC 1150
 QY 249 FSVWIKVPKPVLRVNIAITGVATSECFFPKPGTADKOQSSFCKLCPANSYNSNKGTS 308
 Db 1151 FYADKK-----RLECMQOEG-----CKTCTSNGV-----C 1176
 QY 309 HQC-----DPPDKTSEKGSSCNRPACTKDYFT-----HTAD DANGETQMLY 352
 Db 1177 SECLQWTLNKRDKCVSGSRGCS-----ESEFYSQVEQCRCRHASC----- 1219
 QY 353 KWAKPKICSELEGGAVKLPSAGVKHCPP-----CNPDRFKTNINSTCOPCPY- 399
 Db 1220 -----GSCKNGPADTSCCTSCPPNRLLEOSRCRVSGCREGRFVEAGSLCSCPCLHT 1266
 QY 400 -GSYNSGDCCTRCPAGTEPAVGFEIKWKNWNLPTNMETTVLGSINFEYKGTMGEWAGDH 458
 Db 1267 CSQCVSRTNCNSCSKGLELDQ-----EORTCADCY-VSDRGICAKCYLSCH- 1313
 QY 459 YTAAGASDNDEFMLTUVPGFRPPQSVMDTENKEVARITPVFETLCSVNUCLEYPMGVN 518
 Db 1314 -TCSGPRRNQ-----CVCOPAGHQLAQG 1335
 QY 519 S-RTRNPVPEWFGSKGKOSKVYIIEENTTSFTWAORTTHEASRSKYTDVAKIYSINV 577
 Db 1336 ECHPECPEGFYKSDFGQKCH----- 1357
 QY 578 TNVMNGVASTCRPCALEASDVG-S-CTCSCP-----AGYVYIDRDSGTC 622
 Db 1358 -----YCKTC-----NAGPLACTSCHPSHMSLDDGGLCMECLSSOYYTTSATCKC- 1403
 QY 623 PNTILKAHOPIPGVQAVPGC3GPGTKNNKIHISQYNGC--FSRNPRTTNYNEFALAN 678
 Db 1404 HDSCRCFGP-GQFSKKGCVPPHLQDQLNQCVSQQNQCONQTLAEKSS-----AACNC 1455
 QY 679 -----TVLAGGSPFTSGKLVYIHFHTSLLCG--NOQKMSVCTDNVTDLRIPGES 728
 Db 1456 DGETGECKATASTGGKRTVVGSGSAKSSKHSFPENDNAREFV-----LRIDSP 1509

QY	729	GFSKSITAYVQAOVT-----	--IPPEVYGKACVSSQPVL	762
	:	:	:	
Db	1510	AIT-AIAVACLLITIFSIHAVLQRNSNHVSRSNVRKYKANTSSGRKKNLASAKFTSD	1568	
QY	763	ADRLIGVTDMLTGITSAAEL-----	FHLESGIPDVIFF	798
	:	:	:	
Db	1569	ARFVNIGEDDDTGGDNDSDDELDGNVGTIDINNRIVYDRKGNDHGHEFYIESNDIDAEF	1628	
QY	799	YRSNDVTOCSSGSGSTTRVRCs	821	
	:	:	:	
Db	1629	H-----CNGAGAQRAETQJRCN	1646	

Search completed: March 12, 2003, 00:13:43
Job time : 35 secs

